

# 图像低层特征提取与检索技术

孙君顶 赵 珊 著

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书对基于内容的图像检索技术(CBIR)的基本原理、典型方法和研究进展进行了比较详细的介绍和讨论,并融入了作者多年来的相关研究成果。本书层次分明,内容翔实,理论分析与算法实践相结合,力求实用。本书共7章:第1章介绍CBIR的体系结构、发展历程及发展趋势;第2章介绍CBIR所涉及的关键技术;第3~5章详细讨论了常用图像低层特征(颜色、形状、纹理)的提取与描述算法;第6~7章论述基于压缩域的图像检索技术及高维索引技术。书中各章共列出了400多篇有代表性的参考文献,附在各章的末尾,供读者参考。

本书可作为高等院校计算机科学、信号和信息处理、图书情报等专业的研究生或高年级本科生的专业基础课辅助教材,也可供广大从事模式识别、多媒体分析、信息检索等研究、应用和开发领域的科技工作者和高等院校师生参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

## 图书在版编目(CIP)数据

图像低层特征提取与检索技术 / 孙君顶, 赵珊 著. —北京: 电子工业出版社, 2009.7  
ISBN 978-7-121-08933-6

I. 图… II. ①孙… ②赵… III. 图像数据库—情报检索 IV. G354.49

中国版本图书馆CIP数据核字(2009)第085398号

策划编辑: 董亚峰

责任编辑: 宋兆武 李施诺

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本: 787×1092 1/16 印张: 18.75 字数: 480千字

印 次: 2009年7月第1次印刷

印 数: 2500册 定价: 39.00元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 [zlts@phei.com.cn](mailto:zlts@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

服务热线: (010) 88258888。

# 前言

随着信息社会的发展,大容量存储设备和数字化设备的出现和广泛使用,以及多媒体技术和网络技术的迅速普及,图像数据呈现几何级数的增长趋势,于是出现了大容量的图像数据库。一般的文字搜索引擎对于图像数据来说,几乎束手无策,于是如何迅速、准确地从浩瀚的图像库中检索到所需要的图像成了近年来多媒体领域研究的热点问题。因此,如何将图像处理、计算机视觉等技术与传统数据库技术相结合,建立基于对多媒体内容自动或半自动的新一代有效检索手段便成为迫切的需求。目前,国际上广泛开展的基于内容的图像检索技术(Content-Based Image Retrieval, CBIR)研究就是针对这一挑战的。

针对 CBIR 技术的研究,国外至今已经出版了大量的著作,国际上每年也都召开许多信息处理方面的学术会议,许多大会都有图像检索技术的主题和分会。但是,目前国内有关 CBIR 方面的著作却很少。早期著作如 2002 年清华大学出版社出版的庄越挺等的著作《网上多媒体信息分析与检索》及 2003 年科学出版社出版的章毓晋的著作《基于内容的视觉信息检索》,这两部著作对 CBIR 技术的研究内容作了一定的阐述。2007 年,清华大学出版社出版了周明全等的著作《基于内容图像检索技术》,该书对 CBIR 技术在一定程度上作了较全面的论述。

CBIR 技术涉及图像处理、计算机视觉、机器学习、模式识别、数据库技术及人机交互等诸多领域,内容十分广泛。并且, CBIR 中的图像“内容”包含三个层次:①感知层——视觉上图像的颜色、纹理、形状、轮廓等特征,这些特征属于图像的低层特征;②认知层——图像中的主体、对象以及对象间的关系等;③情感层——用户对图像内容的理解及个人的情感因素。通常认知层及情感层特征又称语义特征。因此,仅靠一本著作来对图像检索技术进行全面的论述是很难完成的,同时也很难在各个方面展开深入的讨论。本书在论述 CBIR 技术时,重点放在图像低层特征(感知层特征)的提取及检索技术上。

全书共 7 章,分为三部分内容。第一部分概述(第 1、2 章):第 1 章,概括介绍图像检索技术的发展、国内外研究现状、应用领域及经典的 CBIR 系统;第 2 章,介绍 CBIR 的基本检索原理、图像内容层次、匹配技术以及性能评价准则等 CBIR 关键技术。第二部分为图像低层特征提取技术(第 3~6 章),详细论述了图像颜色、形状、纹理及压缩域等低层特征提取算法,以及作者在这些领域所取得的研究成果。第三部分(第 7 章)论述了高维索引技术,分析了 CBIR 中“维数灾难”产生的原因,并介绍了常用的解决方法及作者取得的成果。

本书由河南理工大学孙君顶博士及赵珊博士撰写,其中孙君顶撰写第 1~4、7 章及附

录, 赵珊撰写第 5、6 章。本书既参考了许多他人的有关文献, 也汇集了该领域经典成果与前沿研究, 融入了作者近年来在该领域的研究成果。附录的软件原型由研究生张兆生开发。

该书的出版得到河南省基础与前沿技术基金 (072300460050), 河南省“十一五”教育科学规划课题 (2008-JKGHAGH-303)、河南省教育厅自然科学基金 (2008B520012、2009B520013), 河南理工大学博士基金 (B050901, B2009-91) 及河南理工大学骨干教师基金资助。可以说, 没有这些基金的支持, 就不会有书中丰富的研究成果, 也不会有本书的出版。在此, 特致诚挚的谢意!

由于作者水平有限, 书中不妥与疏漏之处在所难免, 敬请读者指正。

编 者  
2009 年 3 月



# 目 录

## 第 1 章 绪论 ..... 1

1.1 图像检索技术发展历程 .....	1
1.1.1 基于文本的图像检索 .....	1
1.1.2 基于内容的图像检索技术 .....	3
1.2 基于内容图像检索技术研究内容 .....	4
1.2.1 特征提取及匹配 .....	4
1.2.2 索引机制 .....	5
1.2.3 用户接口 .....	5
1.3 国内外研究状况 .....	6
1.3.1 国内外研究现状 .....	6
1.3.2 国内外研究热点 .....	7
1.4 CBIR 技术应用 .....	9
1.5 经典 CBIR 系统介绍 .....	10
1.5.1 QBIC .....	10
1.5.2 Virage .....	11
1.5.3 Photobook .....	12
1.5.4 VisualSEEK 和 WebSEEK .....	12
1.5.5 Blobworld .....	14
1.5.6 Netra .....	14
1.5.7 MARS .....	14
1.5.8 SIMPLIcity .....	15
1.5.9 其他系统 .....	17
1.6 本书内容安排 .....	17
参考文献 .....	18

## 第 2 章 基于内容图像检索关键技术 ..... 22

2.1 CBIR 的基本检索原理 .....	22
2.2 图像内容及检索层次 .....	23
2.2.1 图像内容 .....	23
2.2.2 图像检索层次 .....	24
2.3 常用的低层视觉特征描述方法 .....	25
2.3.1 颜色特征 .....	25
2.3.2 纹理特征 .....	27
2.3.3 形状特征 .....	29
2.3.4 MPEG-7 中的图像特征描述符 .....	31

2.4 特征匹配技术 .....	34
2.4.1 计量定理 .....	35

2.4.2 常用的匹配算法 .....	35
2.4.3 精确查询与近似查询 .....	39
2.5 性能评价准则 .....	40
2.5.1 精确度和检索率 .....	40
2.5.2 命中准确率 .....	40
2.5.3 排序值评测法 .....	41
2.5.4 ANMRR .....	41
2.5.5 前 $N$ 个结果的正确率与检索率 .....	41

参考文献 .....	42
------------	----

## 第 3 章 基于颜色特征的图像检索 ..... 49

3.1 引言 .....	49
3.2 颜色空间 .....	49
3.2.1 颜色基础 .....	49
3.2.2 RGB 颜色空间 .....	50
3.2.3 HSV 颜色空间 .....	51
3.2.4 CIEL <sup>*</sup> a <sup>*</sup> b <sup>*</sup> 和 CIEL <sup>*</sup> u <sup>*</sup> v <sup>*</sup> 颜色 空间 .....	53
3.2.5 YCrCb 颜色空间 .....	54
3.3 颜色量化 .....	54
3.3.1 颜色量化的定义 .....	54
3.3.2 常用的颜色量化方法 .....	55
3.4 全局颜色特征 .....	57
3.4.1 颜色直方图 .....	57
3.4.2 改进的颜色直方图方法 .....	58
3.4.3 颜色不变量 .....	58
3.4.4 图像主色 .....	59
3.4.5 色调直方图 .....	59
3.4.6 颜色矢量角直方图 .....	60
3.4.7 颜色矩 .....	63
3.4.8 颜色熵 .....	63
3.4.9 改进的颜色熵及颜色矩 .....	63
3.5 空间颜色特征 .....	69
3.5.1 改进的颜色直方图法 .....	69
3.5.2 颜色聚合向量 .....	70
3.5.3 颜色相关图 .....	70

3.5.4 局部颜色特征 .....	70	4.5.3 基于状态相关图的特征提取 方法 .....	104
3.6 颜色空间分布熵 .....	71	4.6 基于平坦度及凹凸度的描述方法 .....	106
3.6.1 环形颜色直方图 .....	72	4.6.1 平坦度及凹凸度定义 .....	106
3.6.2 改进的环形颜色直方图 .....	72	4.6.2 形状特征量化 .....	107
3.6.3 空间分布熵 .....	73	4.7 基于信息熵的描述方法 .....	110
3.6.4 加权颜色空间分布熵 .....	73	4.7.1 图像信息熵定义 .....	110
3.6.5 消除孤立分布小颜色块的影响 ..	73	4.7.2 图像的单元熵 .....	111
3.7 位平面熵 .....	74	4.7.3 利用熵矩阵的特征值向量进行 检索 .....	113
3.7.1 位平面分解与位平面 .....	74	4.7.4 利用熵矩阵的不变矩进行 检索 .....	114
3.7.2 位平面熵 .....	76	4.7.5 算法特性 .....	114
3.8 位平面熵增强法 .....	77	4.8 基于方向链码的描述方法 .....	121
3.8.1 改进的位平面熵 .....	78	4.8.1 基于链码的形状检索 .....	121
3.8.2 空间分布熵 .....	79	4.8.2 链码分布矢量 (CCDV) .....	123
3.9 基于显著点的图像检索 .....	80	4.8.3 链码相关矢量 (CCCV) .....	124
3.9.1 块逆概率差 (BDIP) 模型及 BDIP 图像的提取 .....	81	4.8.4 链码空间分布熵 (CCSDE) ..	124
3.9.2 显著点提取算法 .....	83	4.8.5 链码相关熵 (CCRE) .....	125
3.9.3 基于显著点的特征提取 .....	84	4.8.6 综合特征描述 .....	127
参考文献 .....	85	4.9 基于角点的描述方法 .....	128
<b>第 4 章 基于形状特征的图像检索 .....</b>	<b>90</b>	4.9.1 轮廓角点提取 .....	128
4.1 形状表达和描述 .....	90	4.9.2 基于内角的轮廓角点检测 .....	129
4.2 基于轮廓的描述方法 .....	91	4.9.3 基于链码局部直方图的角点 检测 .....	129
4.2.1 链码 .....	92	4.9.4 基于 CSS 的角点检测 .....	130
4.2.2 傅里叶形状描述符 .....	92	4.9.5 改进的多尺度角点检测方法 .....	132
4.2.3 曲率尺度空间描述符 .....	93	4.9.6 角点检测算法比较 .....	133
4.2.4 小波描述符 .....	95	4.9.7 距离直方图 .....	135
4.3 基于区域的描述方法 .....	95	4.9.8 相对位置分布 .....	135
4.3.1 几何不变矩 .....	95	4.9.9 相关单元熵 .....	136
4.3.2 Zernike 矩 .....	96	4.10 基于矩的轮廓描述方法 .....	136
4.3.3 ART (Angular Radial Transformation) .....	97	4.10.1 轮廓矩 .....	136
4.3.4 通用傅里叶描述符 .....	98	4.10.2 Chen 不变矩 .....	137
4.4 简单几何参数描述符 .....	98	4.10.3 边界序列矩 .....	138
4.4.1 基于轮廓的方法 .....	98	4.10.4 极半径不变矩 .....	139
4.4.2 基于区域的方法 .....	99	4.10.5 组合矩 .....	139
4.5 基于状态矩阵描述方法 .....	100	参考文献 .....	143
4.5.1 状态矩阵定义 .....	100	<b>第 5 章 基于纹理特征的图像检索 .....</b>	<b>148</b>
4.5.2 基于马尔可夫链形状特征 提取 .....	102		

5.1 图像的纹理描述 .....	148
5.1.1 纹理及纹理特征的定义 .....	148
5.1.2 常用的纹理分析方法 .....	149
5.1.3 纹理的分类 .....	150
5.1.4 纹理研究及应用领域 .....	152
5.1.5 纹理描述存在的问题 .....	153
5.2 统计法纹理分析 .....	153
5.2.1 直方图的矩 .....	154
5.2.2 二维灰度直方图 .....	154
5.2.3 灰度共生矩阵 .....	155
5.2.4 灰度-梯度共生矩阵 .....	159
5.2.5 纹理谱 .....	160
5.2.6 LBP 算法 .....	161
5.2.7 Laws 纹理能量 .....	164
5.2.8 数学形态学分析法 .....	165
5.2.9 自相关函数法 .....	166
5.2.10 Tamura 纹理特征 .....	167
5.2.11 灰度游程长度法 .....	169
5.3 结构法纹理分析 .....	170
5.3.1 结构法基本知识 .....	171
5.3.2 纹理镶嵌 .....	172
5.3.3 Voronoi 多边形方法 .....	173
5.3.4 其他方法 .....	174
5.4 频谱法纹理分析 .....	175
5.4.1 傅里叶变换法 .....	176
5.4.2 贝塞尔-傅里叶变换法 .....	178
5.4.3 小波变换法 .....	179
5.4.4 Gabor 变换法 .....	182
5.5 模型法纹理分析 .....	184
5.5.1 马尔可夫随机场模型法 .....	184
5.5.2 Gibbs 随机场模型法 .....	185
5.5.3 自回归模型 .....	186
5.5.4 多尺度自回归模型 .....	187
5.5.5 基于分形模型 .....	187
5.5.6 Wold 模型法 .....	191
5.6 纹理基元共生矩阵 .....	192
5.6.1 方块编码算法 .....	192
5.6.2 纹理基元的提取 .....	193
5.6.3 纹理基元共生矩阵 .....	196
参考文献 .....	197

## 第 6 章 基于压缩域的图像检索 .....

6.1 概述 .....	201
6.1.1 图像压缩技术 .....	202
6.1.2 静态图像压缩标准 .....	206
6.1.3 压缩域图像检索原理 .....	211
6.1.4 压缩域图像检索的研究内容 .....	213
6.1.5 压缩域图像检索的研究方法 .....	214
6.2 空间压缩域 .....	215
6.2.1 基于矢量量化 .....	215
6.2.2 分形编码 .....	217
6.2.3 预测编码 .....	219
6.3 变换压缩域 .....	220
6.3.1 基于 DFT 变换域 .....	220
6.3.2 基于 DCT 压缩域 .....	221
6.3.3 基于小波压缩域 .....	227
6.3.4 基于 K-L 变换域 .....	231
6.4 空间域和变换域的融合检索 .....	232
6.5 DCT 压缩域内的纹理特征 .....	233
6.5.1 复杂度的定义 .....	234
6.5.2 复杂度直方图 .....	235
6.6 DCT 压缩域内的形状特征 .....	238
6.6.1 理想边缘模型 DCT 块的分类 .....	238
6.6.2 空间边缘分布特征的提取 .....	239
参考文献 .....	241

## 第 7 章 高维索引技术 .....

7.1 高维索引技术 .....	246
7.2 高维索引技术发展趋势 .....	247
7.2.1 向量近似方法 .....	247
7.2.2 近似检索方法 .....	249
7.2.3 并行索引方法 .....	250
7.3 向量空间中的高维特性 .....	251
7.4 维数灾难现象 .....	252
7.4.1 查询代价模型 .....	252
7.4.2 维数灾难现象的产生 .....	253
7.5 基于矢量量化的向量近似方法 .....	254
7.5.1 矢量量化 .....	254
7.5.2 基于矢量量化技术的索引结构 (Cui J.T. et al 2007, 崔江涛 2005) .....	256

7.5.3 码书长度分析与乘积码书法·····	258
7.5.4 近邻搜索算法·····	259
7.5.5 实验分析·····	260
7.6 二次式距离上基于 SVD 的高维索引 方法·····	263
7.6.1 奇异值分解·····	263
7.6.2 索引结构·····	264
7.6.3 近邻搜索算法·····	264
7.7 多分辨率高维索引方法·····	265
7.7.1 基本原理·····	265
7.7.2 索引结构·····	266
7.7.3 近邻搜索算法·····	266
7.8 向量近似方法在相关反馈技术中的 应用·····	267

7.8.1 二次式距离方法·····	268
7.8.2 核函数方法·····	269
7.8.3 改进的近邻搜索算法·····	270
7.9 高维索引技术评价准则·····	271
参考文献·····	272

## 附录 A 基于轮廓的图像检索系统·····277

A.1 系统框架·····	277
A.2 图像数据管理·····	277
A.3 系统实现·····	278

## 附录 B 系统中本文算法实现代码·····288

B.1 基于 MCP 的角点检测算法·····	288
B.2 组合矩算法·····	290

# 第1章 绪 论

随着信息社会的发展，大容量存储设备和数字化设备的出现和广泛使用，以及多媒体技术和网络技术的迅速普及，图像、视频数据呈现几何级数的增长趋势，于是出现了大容量的图像/视频数据库。一般的文字搜索引擎对于图像数据来说，几乎束手无策，于是，如何迅速、准确地从浩瀚的图像数据库中检索到所需要的图像成了近十多年来多媒体领域研究的热点之一。因此，如何将数字图像处理、计算机视觉等技术与传统数据库技术相结合，建立基于对多媒体内容自动或半自动的新一代有效检索手段便成为迫切的需求。近年来，国际上广泛开展的图像检索技术研究就是针对这一挑战的。

本书主要针对基于内容的图像检索技术展开论述，重点讨论图像低层特征提取及检索技术。主要内容包括：基于内容图像检索所涉及的关键技术，图像低层特征（颜色、形状、纹理特征）的提取及描述技术，基于压缩域的图像检索技术和高维索引技术等方面的问题。

## 1.1 图像检索技术发展历程

随着数据库管理技术和计算机视觉的发展，图像检索技术自 20 世纪 70 年代以来，一直成为人们研究的热门课题。从图像检索技术发展的过程来看，主要经历了两个阶段：基于文本的图像检索阶段和基于内容的图像检索阶段（孙君顶 2005，李向阳等 2001）<sup>①</sup>。

### 1.1.1 基于文本的图像检索

基于文本的图像检索技术（Text-Based Image Retrieval, TBIR）的历史可以追溯到 20 世纪 70 年代末期，当时流行的图像检索技术是将图像作为数据库中存储的一个对象，用关键字或自由文本对其进行描述，查询操作是基于该图像的文本描述进行精确匹配或概率匹配，因此这种图像检索技术的实质是采用成熟的文本检索技术来实现对图像的检索。

基于文本的图像检索技术的广泛应用，如 google、baidu、sohu、yahoo 等搜索引擎目前均采用这种方式来检索图像。图 1.1 和图 1.2 分别给出了 baidu 和 google 两个搜索引擎的一次检索结果示例，搜索的文本为“鲜花”，从这两大搜索引擎的检索结果中也可了解该搜索技术的特点。

然而，基于文本的图像检索技术需要人工提前对图像库中的图像进行归纳和注释，图像检索结果也完全依赖于人工标注的图像名称、编号、注释等信息。该技术尚存在以下几个目前无法解决的问题（李向阳等 2001，孙君顶 2005）。

---

<sup>①</sup>括号中内容为参考文献的内容概要。全书下同。

(1) 每一幅图像都需要人工进行注释, 因此标注较大的图像数据库需要大量的人工劳动, 尤其是当资源的数量达到海量级, 或是处理的速度要求接近实时的时候, 完全采用人工的方法都会遇到难以克服的困难。



图 1.1 baidu 针对“鲜花”的搜索结果

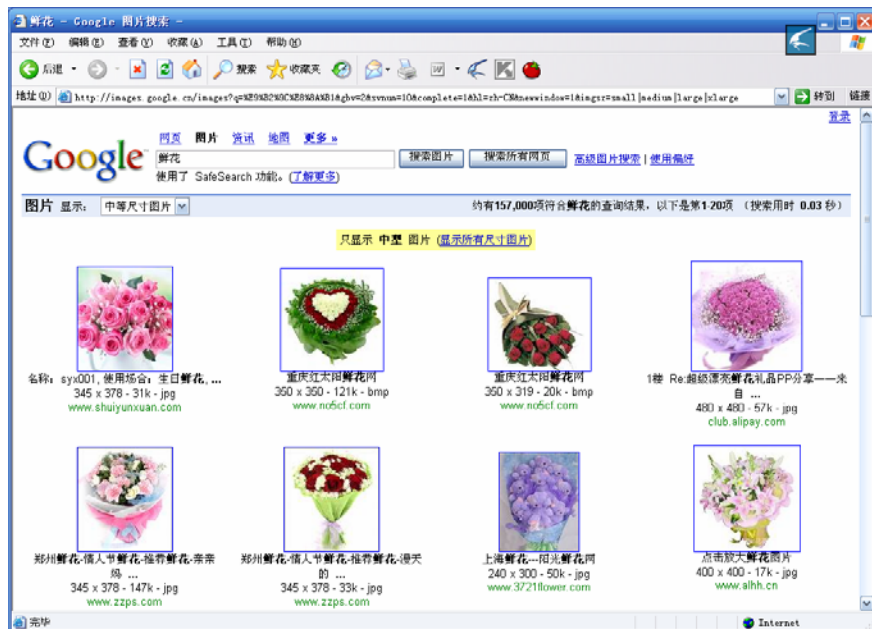


图 1.2 google 针对“鲜花”的搜索结果

(2) 人工注释具有很强的主观性, 对于同一幅图像, 不同人可能有着不同的看法, 因此标注信息会不尽一致; 而且, 一旦人工注释完成后就很难更新和改变, 这使得在很多情

况下文本标注并不能满足实际需求。

(3) 一幅图像所包含的意义非常丰富,“百闻不如一见”、“一图值千言”都说明了这个事实,而人工注释的少量文字很难充分表达一幅图像的内涵。

(4) 由于媒体信息往往是发布在网络环境中,不同国家、不同民族很难用同一种语言对图像进行加注标识,而且对图像语义理解的差异也很大,因此不可能形成一种统一检索方法。

20 世纪 80 年代,虽然多媒体技术发展迅速,图像的获取、创作、压缩、存储技术都取得了举世瞩目的成就,但对图像信息的管理尚未给予足够的重视。90 年代初期,随着大规模数字图像库的出现,上述的问题变得越来越突出。为了实现自动化、智能化、人性化的图像查询和管理方式,使查询者可以实现方便、快速、准确的查找,管理者可以从大量的单调的人工管理工作中解放出来,实现管理工作的无人干预,一种新的图像检索技术——基于内容的图像检索技术被提出并迅速发展起来。

### 1.1.2 基于内容的图像检索技术

基于内容的图像检索,属于图像分析的一个研究领域。它的英文术语叫做CBIR(Content-Based Image Retrieval)。基于内容的图像检索指的是查询条件本身就是一幅图像(或者是对于图像内容的描述),它建立索引的方式是通过提取图像特征,然后通过计算比较这些特征和查询条件之间的距离,来决定两幅图像的相似程度。为了提高检索性能,还需要有效的索引技术以满足检索的实时性要求。这里,图像内容的描述及提取不再依赖于人的手工标注,而是借助于从图像中自动提取的视觉特征,检索过程也不再是关键字匹配,而是视觉特征间的相似匹配。

基于内容的图像检索技术的研究涉及到人工智能、计算机视觉、信号处理、模式识别、认知心理学、数据库、人机交互等诸多学科领域,具有重要的理论意义。同时,它也是理论和实践紧密结合的一项研究,其最终目标就是帮助人们更方便、更快捷和更准确地找到所需要的多媒体资源,因此该技术具有巨大的应用前景。

CBIR 的一般框架如图 1.3 所示。

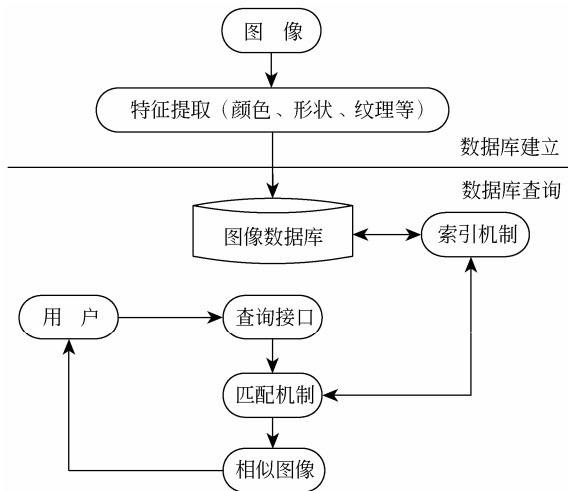


图 1.3 CBIR 一般框架

基于内容的图像检索技术具有以下几个特点 (Berman A. et al 1997, 徐旭 1999):

(1) 直接从图像内容中提取信息线索。基于内容的图像检索, 突破了传统的基于表达式检索的局限, 它直接对图像进行分析和提取特征, 并利用这些描述图像内容的特征来建立索引。

(2) 基于内容的图像检索实质上是一种近似匹配的技术。在数据库中, 需使用模式识别的方法对图像库中的图像按不同索引特征分类; 在检索的过程中, 它采用某种相似度量对图像库中的图像进行匹配, 以获得查询结果。这一点与常规数据库检索的精确匹配方法有明显不同。

(3) 特征提取和索引建立可由计算机自动实现, 避免了人工描述的主观性, 也大大减少了工作量。

若要进一步的了解基于内容图像检索技术的发展及现状, 可参阅文献 (Rui Y. et al 1999, Arnold W.M. et al 2000, Ritendra D. et al 2008), 这些文献分别对基于内容的图像检索在不同阶段的现状及发展趋势做了详细的介绍。

## 1.2 基于内容图像检索技术研究内容

目前, 有关 CBIR 的研究内容很多, 但归纳起来, 主要集中在三个方面: 图像特征提取及匹配、索引机制和用户接口 (Zhang D. S 2002, 孙君顶 2005)。

### 1.2.1 特征提取及匹配

所谓特征提取, 就是从图像中把那些图像自身的内容信息提取出来, 使用户可以据此进行图像检索。特征提取是 CBIR 研究的核心内容, 大量的研究工作都是围绕这个主题开展的。早期的许多研究工作企图找出一组“最佳”特征表示, 它能够最有效率地从图像中提取出来, 并且能够使图像查询的性能变得更好。遗憾的是到目前为止, 并没有哪一种图像表示方法被证明适用于所有应用。因此, 对于不同性质的图像, 需要采用实验方法来确定最适合的图像表示方法。在考察具体的特征提取技术时, 可以从三个方面来了解: 特征提取 (Extraction Method)、特征表示 (Feature Representation) 以及相似度量 (Similarity Measure) (曹奎 2002)。国际标准 MPEG-7 的制定, 其目的是希望使多媒体数据的特征表示标准化, 从而使那些采用 MPEG-7 作为特征表示的系统达到跨系统检索的能力。

图像的特征主要包括低层特征 (Primitive Features) 和语义特征 (Semantic Features) (也称为高层特征)。低层特征主要包括图像的颜色、形状、纹理和空间关系等一些定量的特征, 这些特征可以通过计算机自动或人机交互的方法来提取。图像的语义特征是一种定性特征, 是对图像内容的抽象描述, 语义特征主要通过人工或人机交互 (如相关反馈、机器学习等方法) 的方式提取。而且, CBIR 在不同的应用领域可以采用不同的特征或特征组合进行检索。例如, 在卫星系统中图像的纹理特征是十分重要的, 而在商标配准系统中, 商标的形状特征是最为重要的。

在提取完图像的特征后, 图像检索的主要任务就变成度量图像特征间的相似度问题。合理的相似度量方法也是执行有效图像检索的关键, 常用的相似度量方法主要包括: 欧氏距离、城区距离、二次式距离、直方图相交法等。不同的相似度量方法也有其内在的优点和缺点, 并非某一种度量方法对所有的图像检索系统均适用, 它们也有各自的适用



范围。图像检索时,应根据所提取特征的特点,选择合适的相似度度量方法(洪安祥 2003)。

### 1.2.2 索引机制

索引技术是加速图像相似性检索的关键技术之一,也是多媒体和数据库领域的研究热点和难点。针对多维数据索引机制,研究者提出了许多有效的方法(Lu G.J. 2002, Yu C. 2002),但这些索引结构每种都有自己的适用范围。由于多种因素的影响(如数据的类型、分布情况、维数的大小等),那些适用所有场合、所有数据分布情况的索引结构是不存在的(Bohm C. et al 2001, Chavez E. et al 2001, 董道国等 2002, 刘芳洁等 2003)。

为了有效的完成图像检索,必须解决图像特征的存储格式问题。对于图像的特征,它们之间可能没有内在的顺序,也可能具有多重相关特征,因此在图像检索系统中选用合适的数据结构模型对于有效的图像检索是十分必要的。目前常用的数据结构模型有:R 树(Guttman A. 1984),  $R^+$ 树(Sellis T. et al 1987),  $R^*$ 树(Beckmann N. et al 1990), X 树(Berchtold S. et al 1996), SS 树(White D.A. et al 1996)、SR 树(Katayama N. et al 1997), k-d-B 树(Robinson J.T. 1987)、四叉树(Finkel R. et al 1974)等。在这些数据结构模型中,每一种数据结构模型都有其内在的优点和缺陷,因此在选择合适的数据结构模型时需要考虑所提取的图像特征的特性。

上述各类索引方法在图像特征维数较少的情况下具有很好的检索性能,但后来的研究和实验却发现,这些索引方法在维数足够高的情况下,检索性能下降很快,会退化到顺序查找方法,导致“维数灾难”现象。当图像集的规模变得越来越大时,检索速度就会成为检索系统中的性能瓶颈,这时就需要采用高维数据索引方法来加速检索过程(崔江涛等 2005, 崔江涛 2006)。

### 1.2.3 用户接口

图像检索系统的最终结果是交给用户鉴别的,因此在图像检索系统中,用户接口也起着重要作用,它在用户和检索系统之间提供了一种交互式的检索机制。用户可以通过该接口选取合适的查询机制以及浏览图像检索结果,检索系统也可以根据用户的反馈结果进行学习,进一步提高系统的检索性能。一个理想的视觉查询的描述方式,不仅要方便于用户使用,精确地体现出用户的意图,同时又要包含足够的易于计算机接受的描述性信息。

用户接口提供的常用检索方式主要包括以下几种方式(Boujemaa N. et al 2003, Brunelli R. et al 2000, Egenhofer M. 1996, Chang N.F. et al 1980, 章毓晋 2003)。

(1) 草图查询(Query by Sketching):系统提供一个可以画草图的窗口,用户将想要查找的图像以草图的形式画出来并染上相应的颜色,系统从中提取特征并进行检索。这种方式能提供给用户更大的想象和发挥空间,IBM 公司研制的图像和动态景象检索系统 QBIC(Query by Image Content)就提供了这种查询方式。

(2) 例子查询(Query by Example):这是 CBIR 中最常用的查询方式,就是由用户提供检索示例图像,如通过扫描仪输入图像或由用户从系统中选择查询图像,查询系统根据示例图像自动提取其特征,然后在图像库中找出与示例图像相似的图像并反馈给用户。该方式为用户提供了一种简便的方式来表达图像的内容。

(3) 类别浏览(Browsing by Categories):当用户对要查找的图像比较含糊或用户不

熟悉需要查询图像的具体内容时，可以先按系统的分类体系浏览图像库，待发现感兴趣的目标后再做示例查询。

(4) 图像特征和权值选取 (Feature Selection and Weighting): 用户根据自身的检索要求, 选择图像特征及不同特征的重要性 (权值), 作为检索依据。这些特征包括颜色、纹理、形状、轮廓、空间关系等, 用户在选择特征时, 可以选择单个的也可以是复合的特征, 同时各种特征还可以附加不同的权值。例如, “查找蓝色占 50%, 红色占 50% 的图片”。这种方式在 IBM 的 QBIC 系统里得到了较好的体现。

(5) 相关反馈 (Relevance Feedback) 与机器学习 (Machine Learning): 用户根据需求, 首先对系统的检索结果按照系统的提示进行标识 (如由用户标识检索结果中相似或不相似的图像), 系统根据用户标识信息, 通过系统反馈及学习, 进一步返回满足用户要求的图像。

## 1.3 国内外研究状况

### 1.3.1 国内外研究现状

基于内容的图像检索技术始于 20 世纪 90 年代初期, 由于此项技术涉及的领域很多, 以及该技术对信息检索领域的重要作用, 因此迅速成为研究热点。各大研究机构、公司和高校, 如 IBM、MIT、Columbia、Stanford、UIUC 等都对该技术进行了深入研究, 并推出了各自的图像检索系统。各大著名杂志如: IEEE Trans. on PAMI、IEEE Trans. on Image Processing、IEEE Trans. on Multimedia、IEEE Trans. on CSVT、Computer Vision and Image Understanding、Pattern Recognition、Pattern Recognition Letters 等纷纷设专刊介绍该领域研究的成果, 著名的国际会议如: IEEE Conference on CVPR、IEEE International Conference on ASSP、ACM Conference on Multimedia、SPIE Conference on Electronic Imaging 等纷纷设专题交流最新的研究成果 (Ritendra D. et al 2008, 孙君顶 2005)。

在国内, 各大重点院校和科研机构都展开了关于基于内容的图像、视频、音频的检索方法的研究。清华大学结合国家 863 高技术研究发展项目, 开展了“Web 上基于内容的图像检索”的研究, 研究目标是开发能在 Internet/Intranet 环境下, 通过友好的人机界面, 利用主颜色、纹理、颜色分布和轮廓等图像特征或样本图像检索的方法和工具。中国科学院计算技术研究所和北京图书馆联合开发了“基于特征的多媒体信息检索系统 MIRES” (<http://www.intsci.ac.cn/cbir/>), 浙江大学开发了 Photo Navigator、Photo Engine 和 WebscopeCBR 等系统。总之, 目前越来越多的科研机构都加入了研究基于内容图像检索技术的行列中来。

目前, CBIR 是一个非常繁荣的研究与应用领域, 几乎每天都有新技术报道。随着该领域的逐步发展, 研究者与开发者所面临的问题也相应地发生了变化。大部分的前期研究主要关心如何从一个大图像集中利用自动导出的图像特征检索出图像; 目前的研究工作则集中在如何改进 CBIR 技术上, 以研制出能满足用户实际需求 CBIR 系统, 研究的主要问题包括新类型的图像特征、表示方法以及相似度度量等。显然, CBIR 系统不可能只用传统的文字查询方式达到目的, 但这并不意味着不能将文字查询包含在 CBIR 系统中。事实上, 已经有一些研究工作将基于文本的检索与 CBIR 集成在一起, 即将图像的文字描述与图像的低层特征集成在一个 CBIR 系统中。事实上, 研究人员已经发现全自动的特征提取

似乎是不可能的,而使用者有时也无法明确地描述其中的一些图像特征。越来越多的研究工作已转向“半自动”的特征提取以及交互式查询方式,现有的交互式 CBIR 系统广泛采用了相关反馈技术及机器学习技术。另一方面,一些研究者逐渐意识到“图像相似”是一个与人的视觉感知有关的概念,然而现有的 CBIR 技术无法完整地建立图像的低层特征与语义信息之间的联系。因此,就提出了“语义图像检索”技术,目前针对该技术的研究所取得了大量的研究成果。考察 CBIR 技术的发展过程不难发现,CBIR 技术的发展已经历了几个发展阶段,每个阶段均体现了人们对“图像相似”概念理解的变化,并且均得到相应技术的支持(包括数据库技术、信息检索、计算机视觉、模式识别等)。

### 1.3.2 国内外研究热点

目前,CBIR 研究主要集中在以下几个方面。(孙君顶 2005,赵珊 2007,Ritendra D. et al 2008)

#### 1. 图像低层特征提取和描述技术

最初的图像检索研究主要集中在如何选择合适的图像全局特征去描述图像内容和采用什么样的图像度量方法进行图像匹配。尽管目前仍然无法有效地从图像低层特征获取高层语义,但是,如何提取一些有效的图像特征描述子以便于检索和存储却有着实际的研究价值。MPEG-7 标准中的视觉特征描述子部分(visual descriptor)就是专门针对图像低层特征(如颜色、纹理和形状)进行研究的,目前许多研究团体和研究机构也正从事这方面的研究。

#### 2. 基于区域的图像检索(Region-Based Image Retrieval)

基于区域的图像检索的主要思想是通过图像分割技术(包括自动及半自动方法)将图像划分为不同的区域,然后对于每一个区域使用局部特征来进行描述,综合区域的局部特征从而得到图像的总体特征描述,最后使用合适的相似度度量标准来检索图像。由于该方法更加贴近于用户查询时的思路,同时可以看做是由图像低层特征到图像语义特征的一个过渡,因此成为目前图像检索技术的一个重要研究方向。但自动图像分割是一个相当困难的技术,目前还无法使分割出的区域与图像中的对象很好地对应起来,因此,目前借助于图像分割技术进行基于区域的图像检索的检索准确率并不高。但随着图像分割技术的发展,以及通过人工参与的半自动的分割方式,基于区域的图像检索无疑是一种非常有效的图像检索方法。

#### 3. 图像语义特征提取

基于内容的图像检索技术所指的内容主要包括图像的颜色、形状、纹理和语义等特征。其中,图像的颜色、形状和纹理等特征具有相对直观的特点,而语义特征具有相对主观抽象的特点。就图像检索技术的本身来说,基于语义特征的图像检索方式是最合理的图像检索方式,也最符合用户的需求。

语义检索是“智能”CBIR 系统的主要标志,图像语义检索技术的研究与发展可能是 CBIR 走向成熟与实用的关键。语义图像检索技术的研究中要大量借鉴相关领域的研究成果,例如人工智能、信息检索、认知科学等学科领域,并期待研究出新的知识表示方法以建立图像的低层特征与高层特征(概念描述)之间的联系。人工智能在语义特征的表示方面已经有许多成果,可以应用在 CBIR 中分析图像视觉特征和图像语义特征之间的映射关系,使得高层概念和低层视觉特征之间沟通成为可能,以缩短人机之间对相似图像理解的

差距。此外,目前国际标准化组织制定的 MPEG-7 标准,其目标就是实现集高层语义特征和低层视觉特征于一体的基于内容的多特征综合检索。在研究语义图像检索技术时,需要采用认知科学的研究成果来分析图像内容的特征和人对图像的认知。图像信息在人脑中的长期记忆为心像,人对心像的记忆、检索等操作过程实际上是形象思维过程,因此形象思维科学中关于心像的表征和计算模型将对 CBIR 提供一定的指导。

但目前这种检索方式还面临三个方面的问题(王惠锋等 2002):

(1) 必须提供图像语义的有效描述方式;

(2) 必须有提取图像语义描述的方法;

(3) 语义检索系统的语义处理方法。因此,虽然人们更偏爱语义查询,但是这种查询方式的完全智能化目前还较难实现。目前基于语义特征的图像检索主要研究如何从多种渠道获取图像语义信息、所获取的语义信息如何与图像低层特征结合并通过相关反馈在图像之间传递语义信息,以及如何将图像低层特征与图像的关键字结合进行图像的自动标注以提高检索的准确率等(Lu Y. et al 2000, Zhu X.Q. et al 2000, Han J.W. et al 2002, Zhao R et al 2002a, Zhao R. et al 2002b)。

#### 4. 高维索引技术

网络的飞速发展导致产生了大型的图像数据库,但最新的研究模型也只能处理几百或几千幅图像,因为只有这样,在顺序扫描处理这些图像时才不至于严重影响系统的操作性。多维索引方法在特征维数较少的情况下具有很好的检索性能,但在维数足够高的情况下,检索性能下降很快,其效率甚至会低于最原始的顺序查找方法。这一现象已经在几乎所有的传统索引方法中得到印证(崔江涛 2005)。随着图像数量的日益增多,检索速度已经成为瓶颈。尽管在这一研究领域已取得一些进展(如 1.2.2 节提到的多种树型索引结构),但探索更加有效的高维索引技术仍是一个急需解决的问题。

#### 5. 相关反馈技术

“计算机视觉”、“模式识别”与 CBIR 的基本区别就在于人在系统中的作用不同,前者依赖于计算机的处理能力企图达到“全自动”,而后者加入了人的协同工作,是一种“半自动”。在 CBIR 系统中,人是不可缺少的一个组成部分,因此需要探索人与机器的协同工作。早期研究侧重于“完全自动化的系统”,并且试图寻找一种“最佳图像特征”,然而这样的思路并没有获得成功的 CBIR 系统。越来越多的研究工作则着重于交互式系统并且强调人在系统中的作用。现有的交互式 CBIR 系统主要采用“相关反馈”(Relevance Feedback)技术。相关反馈技术主要基于人机交互的思想,借助一种相关反馈的技术来猜测用户的需求,并且根据用户的需求动态调整系统检索时所采用的特征向量或参与检索的不同特征的权重系数,从而尽量缩小低层特征和高层语义之间的差距,提高算法的检索效果。其实,相关反馈是文本检索领域中一个基本的技术,Rui Yong 最先将其用到 CBIR 领域(Rui Y. et al 1998, Rui Y. et al 2000),实验证明它十分有效。近年来,将反馈技术用于图像检索中也是目前图像检索技术研究的一大热点,许多反馈方法被提了出来并用于图像检索中,有关反馈算法的综述可参见(吴洪 2004, 吴洪等 2005, 谢晓霞 2006)。目前,相关反馈技术中的主要的问题包括人机交互方式、用户模型、相关判断、学习算法等。

#### 6. 相关反馈与机器学习结合技术

由于图像检索系统的最终用户是人,因此通过交互手段来捕获人对图像内容的认知是相当重要的。为了把用户模型嵌入到图像检索系统中,最近几年在基于内容的图像检索领

域引入了相关反馈与机器学习 (Machine Learning) 机制, 将成熟的学习算法与图像检索中的在线学习过程 (On-line Learning) 结合起来以提高检索准确率。代表性的工作包括: 基于 Bayesian 理论 (Cox I.J. et al 2000), 基于 SVM (Support vector machine) (Tong S. et al 2001), 基于 Active Learning (Zhang C. et al 2002), 以及基于其他的学习理论 (Qian F. et al 2002, Wang J.Z. et al 2002) 等方法。

## 7. 性能评价

目前基于内容的图像检索技术的性能评价主要借鉴了文本检索中的一些评价方法, 由于图像内容的主观性使得对图像检索效果的评价仍没有通用的测试图像集和公认的评价标准, 并且缺少在相同数据集和查询条件下对不同检索系统进行有效性的对比实验, 也难以确定什么样的评价标准更为有效, 因此检索性能的评价还处在很不成熟的阶段, 需要进一步研究。

对检索系统的合理评价, 需要有一套能够平衡表达各种场景和事物的标准测试数据来评价检索的效率和效果, 就像在图像处理领域, 大家都用Lena图像作为实验图像一样 (Vogel, J. et al 2006, Wong Y.M. 2007, Liu W.Y. et al 2001)。当然, 性能评价是一项复杂的工作, 要召集该领域专家以及收集大量有代表意义的图像数据, 以便能够测试各种算法的效率, 并在此基础上定义标准的性能的评价准则, 这样才可以利用标准的检索性能评价准则来全面的评价检索算法的性能。

# 1.4 CBIR技术应用

CBIR 技术将对大规模图像信息的管理和访问提供有力支持。它可以广泛应用于信息检索服务、犯罪预防、医疗诊断、新闻和广告、商标和知识产权、地理信息和远程遥感、教育培训和军事等领域, 目前比较成熟的应用有指纹识别、人脸识别和图像搜索引擎等 (Gudivada V. N. et al. 1995, Eakins J. P. et al. 1999, Rui Y. et al. 1999)。

## 1. 信息检索

面对日益增长的图像信息和图像检索需求, 传统的检索手段显得愈发笨拙和不适宜, 而 CBIR 技术可以自动对图像库做基于内容的索引, 查找出与检索图像相似的图像, 更好的满足基于内容的检索需求。

## 2. 知识产权

科技的飞速发展使得人们越来越关注知识产权的保护问题, 而许多知识产权的载体都是图像。如一个新商标在注册前需要和已经注册的商标做比较以确定它是否与其他商标雷同, 以免造成侵权行为, 这就需要计算机对庞大的商标图像库进行视觉相似性检索, 直接比较其颜色或形状特征来确定是否相似。

## 3. 犯罪预防

通过指纹、鞋印、人脸来确认和查找凶犯是目前安全部门常用的手段。计算机检索系统根据指纹、鞋印和人脸的图案信息进行基于特征的相似性鉴别, 这种技术已被世界各地安全部门广泛采用。

## 4. 医疗诊断

现代医疗器械和技术的发展产生了大量的医用图像信息, 如 X 光片等。人们发现通过以往相似病例的诊断图像来做辅助诊断是一项非常有意义的事情, 这就需要在图像库里

查找出与当前诊断图像相似的图像。目前,医学图像检索已成为基于内容图像检索技术应用的一个主要方向。

### 5. 教育与培训

在教育与培训领域,例如远程教学、交互式培训、自学教育及雇员再教育等有着广阔的应用前景。国外在培训系统领域已投入了大量的经费,开展了相关课题的研究工作。我国多媒体教学研究工作的开展、网上教学与辅导已进入实用阶段,这些都为图像数据库应用于教育培训领域提供了广阔的前景。

### 6. 数字图书馆

随着现代信息技术革命的深入发展,图书馆正在发生前所未有的变革,正在向数字图书馆迈进。数字图书馆实际上是一个数字信息资源库,其中有字符数值库、文本库、声音库、图像库等。因此,如何快速、高效地从数字图书馆中找出用户所需的信息就成为现代图书馆研究的热点和关键技术之一。

### 7. 工业与商业领域

工业应用包括企业多媒体信息系统、CAD/CAM 等;商业应用有电子商务、在线广告、在线购物、股票等。

### 8. 军事领域

CBIR 技术在军事领域里有着巨大的应用需求。例如,从雷达图像里识别敌机,从卫星照片里识别目标,为巡航导弹提供制导系统等,此外,在犯罪预防和信息检索领域里的应用也可以应用在军事领域中。

总之,CBIR 技术是一项快速发展的并且颇具潜力的前瞻性技术,在许多领域都具有很高的应用价值。

## 1.5 经典CBIR系统介绍

鉴于基于内容图像检索系统的重要性、有效性和优越性,从 20 世纪 90 年代初期,各大公司和科研机构陆续推出了一些商用或研究用的图像检索系统。本节将简单介绍几个经典的软件系统。

### 1.5.1 QBIC

QBIC (Query by Image Content) (Niblack W. et al 1993, Flickner M. et al 1995, Niblack W. et al 1997) (<http://www.qbic.almaden.ibm.com/>) 是由 IBM 公司于 20 世纪 90 年代开发制作的图像和动态景象检索系统,是在基于内容图像检索领域应用最早的商用产品,它的系统框架和结构对后来的图像检索系统具有深远的影响。

QBIC 系统提供了多种查询方式,包括利用标准范图(示例图像)检索、用户绘制简图或扫描输入图像进行检索、选择色彩或结构查询方式、用户输入动态影像片段和前景中运动的对象检索。QBIC 中使用的颜色特征有色彩百分比、色彩位置分布等。使用的纹理特征是根据 Tamura 提出的纹理表示的一种改进(Tamura H. et al 1978),即结合了粗糙度、对比度和方向性的特性。使用的形状特征有面积、圆形度、偏心率、主轴偏向和一组代数矩不变量。另外,QBIC 系统还考虑到了高维特征的索引,采用  $R^*$  树作为索引结构。另外,QBIC 系统还支持基于文本的关键字查询方式。总之,该系统技术成熟,功能全

面，为基于内容的图像检索技术的验证和推广作出了很大贡献。其查询界面如图 1.4 和图 1.5 所示。

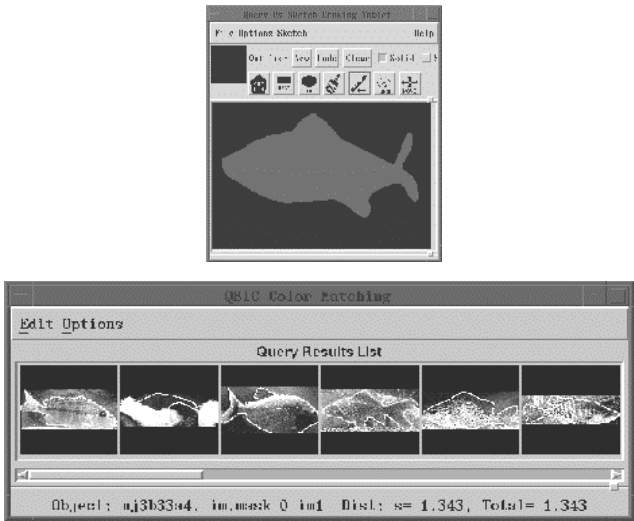


图 1.4 QBIC 基于形状特征检索示例

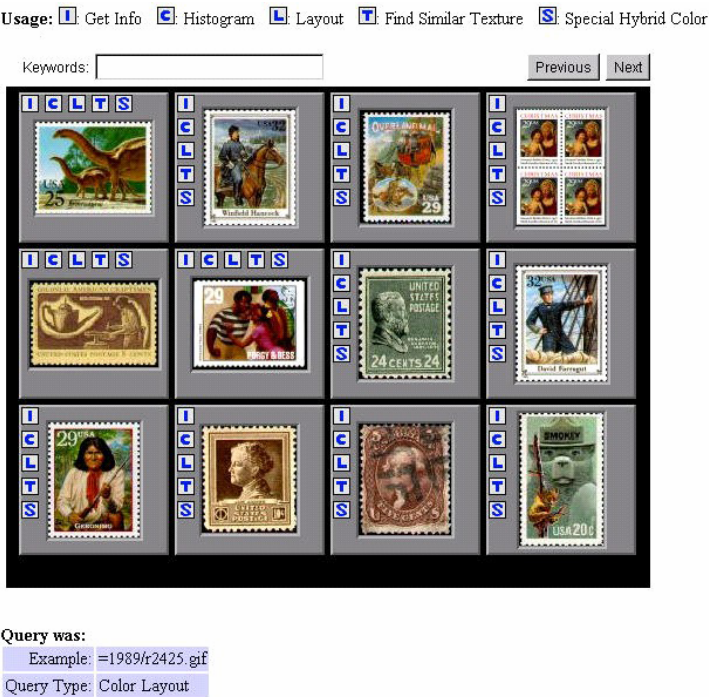


图 1.5 QBIC 系统查询界面

1.5.2 Virage

Virage (Bach J.R. et al 1996) 是由 Virage 公司开发研制的基于内容的图像搜索引擎。与 QBIC 系统相类似，Virage 也支持基于颜色、颜色布局、纹理和结构等视觉特征的图像

检索。同时, Virage 系统还支持上述四个特征的任意组合, 而且用户还可以根据实际的需要对不同的特征赋予相应的权重来改变查询线索的重要性。Virage 技术的核心是 Virage Engine 以及在图像对象层上的操作。它将查询引擎作为一个插件, 既可以应用到通用的图像查询中, 也可对其进行扩展并应用到特定的领域。

### 1.5.3 Photobook

Photobook (Pentland A. et al 1996) 是 MIT 媒体实验室开发的图像检索系统。Photobook 有三个子部分, 分别用于提取形状、纹理和面部特征。图像在装入时按人脸、形状或纹理特性自动分类, 图像根据类别通过显著语义特征压缩编码。因此, 用户可以在这三个子部分中分别进行基于形状、基于纹理和基于面部特征的图像检索。

Photobook 提供了一些交互式工具进行基于内容的浏览和检索, 它使用颜色、纹理、形状、变换域系数建立特征向量, 并提供多种相似性准则, 如欧几里得距离、城区距离、向量内积、小波树距离等。用户亦可以选择一种或几种算法的线性组合计算图像的相似性, 而且还允许用户勾画感兴趣区域, 参与特征提取过程。Photobook 系统的查询界面如图 1.6 所示。



图 1.6 Photobook 系统查询界面 (<http://vismod.media.mit.edu/vismod/demos/facerec/basic.html>)

### 1.5.4 VisualSEEK和WebSEEK

VisualSEEK (Smith J.R. et al 1997a) 和 WebSEEK (Smith J.R. 1997, Smith J.R. et al 1997b) 是由美国哥伦比亚大学开发的姊妹系统。其主要特点是研究利用图像区域空间关系进行查询和从压缩域提取视觉特征来进行检索, 系统中主要使用的特征是颜色特征和基于小波变换的纹理特征。

VisualSEEK 支持基于视觉特征和它们之间空间关系的查询, VisualSEEK 的系统查询界面如图 1.7 所示。WebSEEK 主要是面向 Web 查询的, 采用了先进的特征提取技术, 用



户界面强大，操作简单，查询途径丰富，输出画面生动且支持用户直接下载信息，其查询界面如图 1.8 所示。

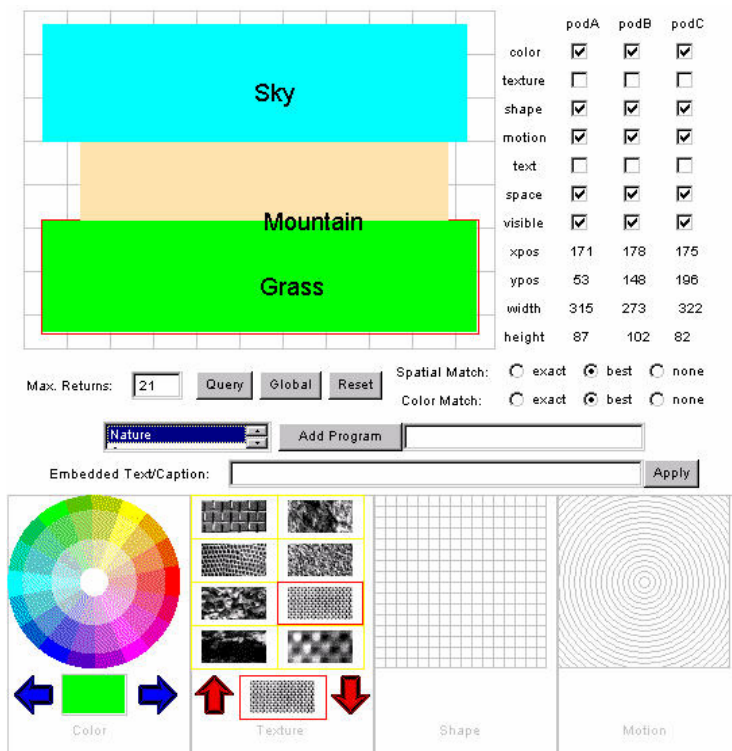


图 1.7 VisualSEEK 系统查询界面

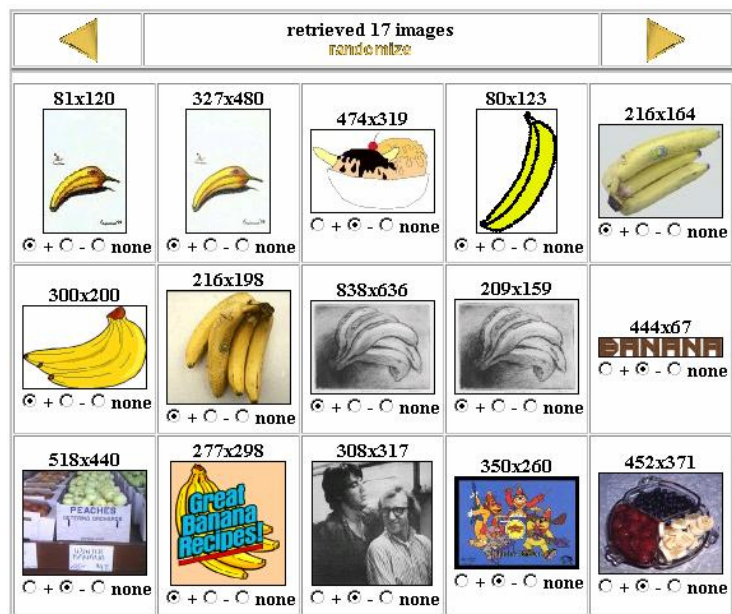


图 1.8 WebSEEK 系统查询界面

### 1.5.5 Blobworld

Blobworld (Carson C. et al 1999, Cllad C. et al 2002) 是由 UC BERKELEY 大学开发的基于区域的图像检索系统。该系统使用的图像特征为颜色、纹理、位置以及区域和背景的形状。在颜色方面使用 Lab 空间的 218bin 的颜色直方图进行描述, 纹理通过区域的平均对比度和各向异性来描述, 区域的形状的描述由面积、离心率和方向性组成。对各个特征的相似性分别采用欧式距离和加权欧式距离进行度量, 最后给出统一的相似性距离。同时, 该系统还对 218bin 的颜色直方图的计算进行降维处理, 并采用  $R^*$  树索引结构以提高检索效率。Blobworld 的查询界面如图 1.9 所示。

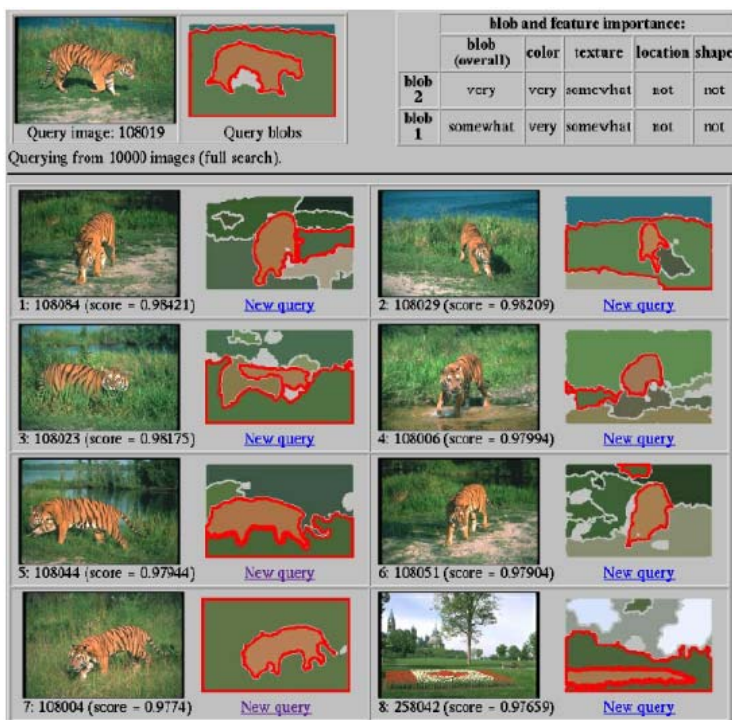


图 1.9 Blobworld 查询界面

### 1.5.6 Netra

Netra (Ma W.Y. et al 1997, Ma W.Y. et al 1999) 是 UCSB 数字图书馆项目中的图像检索原型系统。Netra 系统和其他基于全局特征的图像检索系统不同, 它采用基于区域局部特征的图像检索方案。它首先将图像分割成区域, 然后通过区域的颜色、形状、纹理、空间位置信息表征图像内容。纹理提取采用 Gabor 滤波算子, 区域分割算法基于改进的边缘跟踪算法 (edge flow)。

### 1.5.7 MARS

MARS (Huang T.S. et al 1996, Chakrabarti K. et al 2000, Rui R. et al 1997) 是由 UIUC 开发的支持图像底层特征的复合检索的图像检索系统。它的应用范围相当广泛, 包

括计算机视觉、图像数据库检索和信息检索等多个领域。其特点是使用比较全面的图像底层特征，提供基于树结构的多特征的组合检索。在图像特征方面：使用 HSV 空间的 HS 上的色彩直方图来描述图像的颜色；提取图像纹理的粗糙度和方向性以及对比度等特征描述纹理；采用图像的规则分割的方法对图像特征的空间分布进行描述（颜色直方图和小波变换系数）；根据纹理对图像进行分割来实现图像中的对象描述，并对分割后的对象区域按照敏感性进行分组；使用 Fourier 描述子对图像中对象的形状进行描述。检索时对上述特征分别采用相应的相似度度量方法，最终给出综合排名。由于采用多方面的图像特征描述与相似度度量方法，该系统提供较复杂的检索功能，如可以，则通过布尔表达式进行组合检索。MARS 的焦点不在于找到单一的最佳特征表达，而是如何把不同的视觉特征组织成为一个可以动态适应于不同应用和不同用户的有意义的检索机制。这个系统的突出特点在于引入了相关反馈机制，能够根据用户的交互，动态的组织和优化查询，提高检索效率。其检索效果如图 1.10 所示。

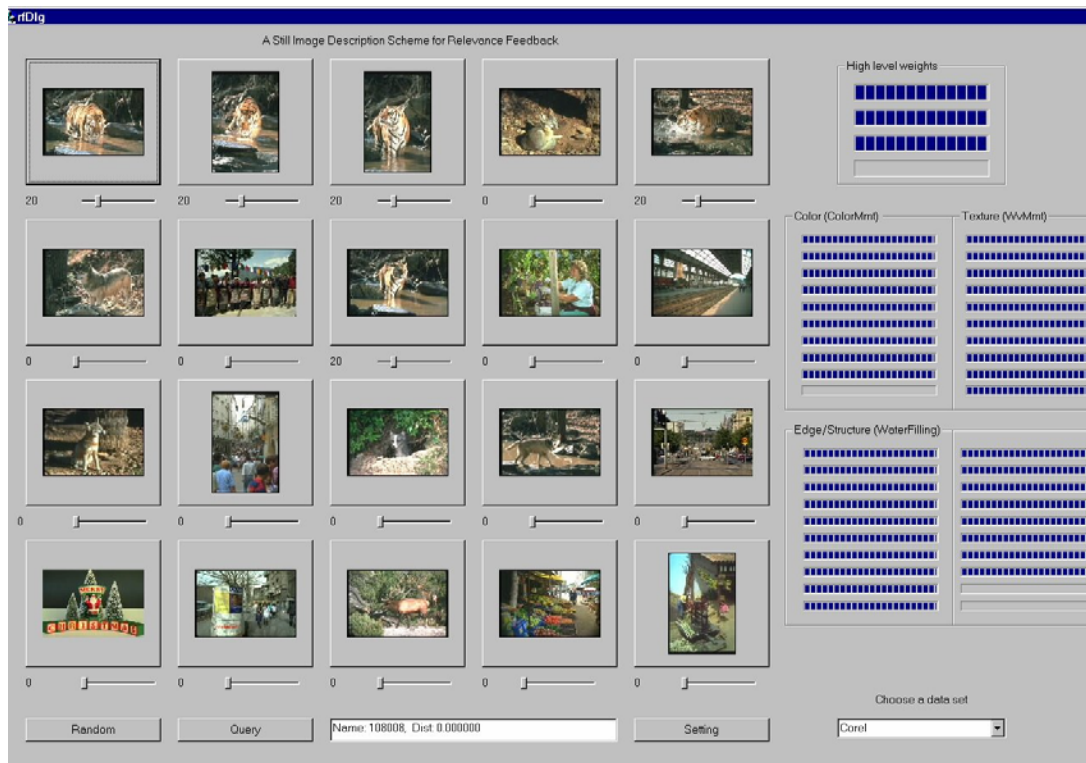


图 1.10 MARS 检索效果图

### 1.5.8 SIMPLicity

SIMPLicity (James Z. Wang et al 2001, Jia Li et al 2003) 是由斯坦福大学与滨州大学开发的。系统提出了一种综合区域匹配 (Integrated Region Matching, IRM) 来定义图像的相似度。它具有旋转和平移不变性，允许区域间的一对多匹配，对不准确的分割比较稳健。图 1.11 给出了其检索界面，图 1.12 给出了其中的一次检索结果。其网址见：[http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch\\_show.cgi](http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi)。

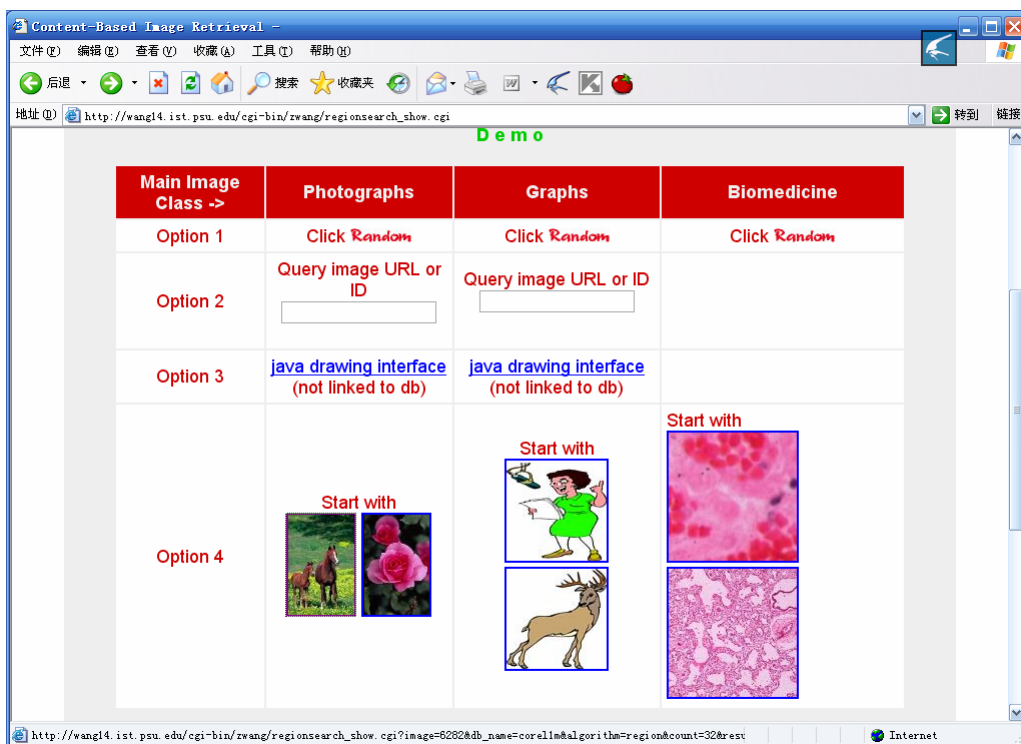


图 1.11 SIMPLicity 系统检索界面

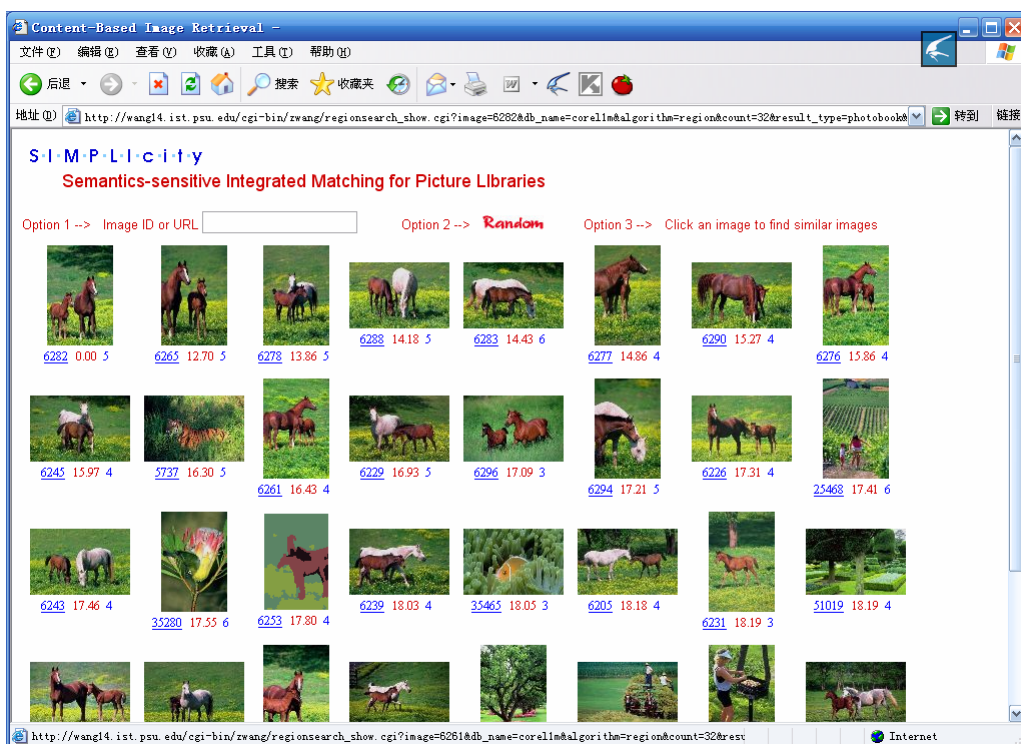


图 1.12 SIMPLicity 一次检索结果



### 1.5.9 其他系统

除了上述介绍的系统外，还有更多的图像检索系统，可参阅文献（Remco C.-V. et al 2000），也可参阅网站<http://www.aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/cbir-survey.html>，里面介绍了 43 个图像检索系统的基本内容。图 1.13、图 1.14 给出了中国科学院计算技术研究所和北京图书馆联合开发的“基于特征的多媒体信息检索系统MIRES”的基本检索界面。



图 1.13 MIRIS 基本检索界面



图 1.14 MIRIS 一次检索结果

## 1.6 本书内容安排

国内外至今已经出版了大量有关基于内容图像检索技术方面的文章，有关图像检索技术的综述文献也已发表了很多，代表性的有（李向阳等 1997，Rui Y. et al 1999，Arnold W.M. et al 2000，Ritendra D. et al 2008）等。国际上每年也都召开许多信息检索的学术会

议,许多大会都有图像检索技术的主题和分会,但有关论述图像检索技术的书籍却很少。早期的著作(章毓晋 2003,庄越挺等 2002)对基于内容图像检索技术作了一定的阐述,2007年,清华大学出版社出版了一部针对基于内容图像检索的专著(周明全等 2007),对基于内容的图像检索技术作了较全面的论述。但对于基于内容的图像检索技术来说,其涉及的内容十分广泛,若面面俱到,就很难对各个方面展开深入论述。鉴于此,本书在讨论基于内容的图像检索技术时,仅仅针对图像低层特征提取及检索技术方面展开论述,深入讨论图像检索技术中涉及的主要技术,即图像颜色、形状、纹理特征的提取及描述技术,基于压缩域的图像检索技术以及高维索引技术几个部分,共包含7章和2个附录。

从各章的主要内容来看,第1章是绪论,概括介绍图像检索技术的发展和现状,国内外的研究现状,基于内容图像检索技术的应用领域以及介绍一些经典的CBIR系统。

第2章介绍基于内容图像检索中涉及的一些关键技术,分别介绍了CBIR的基本检索原理,图像内容的层次、分类及MPEG-7中的图像特征描述符,图像检索中常用的匹配技术以及图像检索系统常用的性能评价准则。

第3章介绍基于颜色特征的图像检索技术,主要比较详细的介绍了图像检索中常用的颜色空间,颜色量化方法,全局颜色特征提取及描述方法,空间颜色特征提取及描述方法,最后介绍了两种新的颜色空间特征描述方法:颜色空间分布熵及位平面分布熵。

第4章介绍基于形状特征的图像检索技术,主要比较详细的介绍了形状表达和描述方法的分类,基于轮廓的经典描述方法,基于区域的经典描述方法。在此基础上,介绍了几种新的形状描述方法:基于状态矩阵的描述方法,基于平坦度和凹凸度的描述方法,基于信息熵的描述方法,基于方向链码的描述方法,基于轮廓角点的描述方法及基于轮廓矩的描述方法。

第5章介绍基于纹理特征的图像检索技术,首先分别介绍了常用的四类纹理描述方法:统计法、结构法、频谱法及模型法,然后介绍了一种基于纹理基元共生矩阵的纹理描述方法。

第6章介绍基于压缩域的图像检索技术,包括空间压缩域和变换压缩域中常用的描述算法,并介绍了两种基于DCT压缩域内图像纹理及形状特征提取方法。

第7章介绍高维索引技术,首先分析了“维数灾难”产生的原因,介绍了目前常用的高维索引方法及评价方法,在此基础上,介绍了基于矢量量化的向量近似方法、二次式距离上基于SVD的高维索引方法及多分辨率高维索引方法。

附录内容是结合基于轮廓特征的检索技术,实现了一个检索系统。

## 参 考 文 献

- [1] Arnold W.M., Marcel W., Simone S., et al, 2000. Content-based image retrieval at the end of the early years. IEEE Trans. On PAMI, 22(12): 1349-1379.
- [2] Bach J.R., Fuller C., Gupta A., et al., 1996. The Virage image search engine: an open framework for image management, In Proc. SPIE Storage and Retrieval for Image and Video Database, 76-87.
- [3] Beckmann N., Kriegel H.P., Schneider R., Seeger B. 1990. The R\*-tree: an efficient and robust access method for points and rectangles. In Proc. ACM SIGMOD Int. Conf. on Management of Data, 322-331.
- [4] Berman A., Shapiro L., 1997. Efficient image retrieval with multiple distance measures, SPIE3022, 12-21.
- [5] Berchtold S., Bohm C., Kriegel H. P. 1996. The X-tree: an index structure method for high-dimensional

- data. In Proc. ICVLDB. 28-39.
- [6] Bohm C, Berchtold S., Keim D.A., 2001. Searching in high-dimensional spaces-indexing structures for improving the performance of multimedia databases. *ACM Computing Surveys*, 33(3): 322-373.
  - [7] Boujemaa N., Fauqueur J., Valérie G., 2003. What's beyond query by example? Research Report (2003) [inria-00071515 – version 1].
  - [8] Brunelli R., Mich O., 2000. Image retrieval by examples, *IEEE Trans. Multimedia*, 2(3): 164-171.
  - [9] 曹奎. 基于颜色的图像检索技术研究. 博士学位论文, 华中科技大学, 2002.
  - [10] Carson C., Thomas M., Belongie S., et al., 1999. Blobworld: a system for region-based image indexing and retrieval. *Proc. Visual Information Systems*, 509-516.
  - [11] Chakrabarti K., Binderberger M.O., Porkaew K., et al, 2000. Similar shape retrieval in MARs, *IEEE Int. Conf. On Multimedia and Expo (II)*. USA, 709-712.
  - [12] Chang N.F., Fu K.S., 1980. Query by pictorial example. *IEEE Trans. on Software Engineering*, SE-6(6): 519-524.
  - [13] Chavez E., Navarro G., Baeza-Yates R., Marroquin J.L., 2001. Searching in metric spaces. *ACM Computing Surveys*, 33(3): 273-321.
  - [14] Cllad C., Serge B., Hayit G., et al., 2002. Blockworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans on PAMI*, 24(8): 1026-1038.
  - [15] Cox I.J., Miller M.L., MinLka T.P, et al, 2000. The Bayesian image retrieval system, PicHunter: theory, implementation and Psychophysical experiments, *IEEE Trans. On Image Processing*, 9(1): 20-37.
  - [16] 崔江涛, 孙君顶, 周利华. 基于小波变换的多分辨率高维图像检索方法. *西安电子科技大学学报*, 2005. 32(3): 370-373.
  - [17] 崔江涛. 高维索引技术中向量近似方法研究. 博士学位论文, 西安电子科技大学, 2005.
  - [18] 董道国, 薛向阳, 罗航哉. 多维数据检索结构回顾. *计算机科学*, 2002, 29(3): 1-6.
  - [19] Eakins J.P., Graham M.E., 1999. Content-based image retrieval: a report to the JISC technology applications programme. Institute for image data research: Newcastle-upon-tyne, 1-63.
  - [20] Egenhofer M., 1996. Spatial-query-by sketch. *IEEE Symposium on Visual Languages*, IEEE Computer Society, 60-67.
  - [21] Finkel R., Bently J.L., 1974. Quad trees: a data structure for retrieval of composite keys. *Acta Inf.* 4(1): 1-9.
  - [22] Flickner M., Sawhney H., Niblack W., et al. 1995. Query by image and video content: the QBIC System. *IEEE Computer*, 28(9): 23-32.
  - [23] Gudivada V.N., Raghavan V.V., 1995. Content-based image retrieval systems. *IEEE Computer*, 28(9): 18-22.
  - [24] Guttman A., 1984. R-trees: A dynamic index structure for spatial searching. *The ACM SIGMOD Int'l Conf. on Management of Data*, Boston, MA, 47-57.
  - [25] Han J.W., Guo L., 2002. A new image retrieval model supporting query by semantics and example. *IEEE Conference on Image Processing*, New York, USA.
  - [26] Huang T.S., Mehrotra S., Ranlchandran K., 1996. Multimedia analysis and retrieval system (MARS) project, In *Proc. of 33rd Annual Clinic on Library Application of Data Processing Digital Image Access and Retrieval*.
  - [27] 洪安祥. 基于内容的图像检索若干论题研究. 博士学位论文, 浙江大学, 2003.
  - [28] James Z. Wang, Jia L., Gio W., 2001. SIMPLicity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9): 947-963.
  - [29] Jia L., James Z. Wang., 2003. Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9): 1075-1088,
  - [30] Katayama N., Satoh S., 1997. The SR-tree: an index structure for high-dimensional nearest neighbor queries. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, Tucson, Arizon USA, 369-380.
  - [31] Liu W.Y., Zhong S., Li S., et al, 2001. A performance evaluation protocol for content-based image retrieval

- Algorithms/Systems. Proc. IEEE CVPR Workshop on Empirical Evaluation in Computer Vision, Kauai, USA.
- [32] Lu G.J., 2002. Techniques and data structures for efficient multimedia retrieval based on similarity. *IEEE Tran. on Multimedia*, 4(3): 372-384.
- [33] Lu Y., Hu C., Zhou X., et al, 2000. A unified framework for semantics and feature based relevance feedback in image retrieval systems. *Proc of ACM Multimedia conference*.
- [34] 刘芳洁, 董道国, 薛向阳. 度量空间中检索高维索引结构回顾. *计算机科学*, 2003, 30(7): 64-68.
- [35] 李向阳, 庄越挺, 潘云鹤. 基于内容的图像检索技术与系统. *计算机研究与发展*, 2001, 38(3): 344-354.
- [36] Ma W.Y., Manjunath B.S., 1997. NETRA: A toolbox for navigating large image database, *Proc. of IEEE International Conference on Image Processing*, Santa Barbara, California, USA, 925-928.
- [37] Ma W.Y., Manjunath B.S., 1999. NETRA: A toolbox for navigating large image databases, *Multimedia Systems*, 7(3): 184-198.
- [38] Niblack W., Barber R., Equitz W., et al, 1993. The QBIC project: querying images by content using color, texture, and shape. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, 2-3 February '93, San Jose, CA, pages 173-187.
- [39] Niblack W., Zhu X., Hafner J.F., et al., 1997. Updates to the QBIC system, *SPIE*, 3312, 150-161.
- [40] Pentland A., Rosalind W., Stanley S., 1996. Photobook: content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3): 233-254.
- [41] Qian F., Li M.J., Zhang L., et al. 2002. Gaussian mixture model for relevance feedback in image retrieval. *IEEE conf. on ICME*.
- [42] Remco C.-V., Mirela T., 2000. Content-based image retrieval systems: A survey, Technical Report UU-CS-2000-34, Department of Computing Science, Utrecht University.
- [43] Ritendra D., Dhiraj J., Li J., et al., 2008. Image retrieval : ideas, influences, and trends of the new age. *ACM Transactions on Computing Survey*, 40(2): 1-66.
- [44] Robinson J.T., 1987. The K-D-B-tree: a search structure for large multidimensional dynamic indexes. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 433-438.
- [45] Rui Y., Huang T.S., Mehrotra S., 1997. Content-Based image retrieval with relevance feedback in MARS. In: *Proc. of the IEEE Int'l. Conf. on Image Processing*, II. IEEE Computer Society. 815-818.
- [46] Rui Y., Huang T S, Mehrotra S., 1998. Relevance feedback: a Powerful tool in interactive content-based image retrieval. *IEEE Trans. On CSVT*, 8(5): 644-655.
- [47] Rui Y., Huang T., Chang F., 1999. Image retrieval: Current techniques, promising directions, and open issues. *Visual Comm. Image Represent*, 10(1): 39-62.
- [48] Rui Y., Huang T. S., 2000. Optimizing learning in image retrieval. *IEEE Conf. On CVPR*, South Carolina, USA.
- [49] Sellis T., Kriegel H.P., Faloutsos C., 1987. The R+-tree: a dynamic index for multidimensional objects. In *Proc. 13th Int. Conf. on Very Large Databases*, Brighton, GB, 507-518.
- [50] Smith J.R., 1997. Integrated spatial and feature image systems: retrieval, compression and analysis. PhD thesis, Graduate School of Arts and Sciences, Columbia University.
- [51] Smith J.R., Chang S.F., 1997a. Visually searching the web for content, *IEEE Multimedia Magazine*, 4(3): 12-20.
- [52] Smith J.R., Chang S.F., 1997b. Querying by color regions using the VisualSEEK content-based visual query system. In M. T. Maybury, editor, *Intelligent Multimedia Information Retrieval*. AAAI Press.
- [53] Su Z., Zhang H.J., 2001. Relevance feedback using bayesian classifier in content-based image retrieval. *Proceedings of SPIE, the International Society for Optical Engineering*, 4315: 97-106.
- [54] 孙君顶. 基于内容的图像检索技术研究. 博士学位论文, 西安电子科技大学, 2005.



- [55] Tamura H., Mori S., Yamawaki T., 1978. Texture features corresponding to visual perception. *IEEE Trans. on Systems, Man, and Cybernetics*, 8(6): 460-473.
- [56] Tong S., Chang E., 2001. Support vector machine active learning for image retrieval. *Proc. of ACM Multimedia*, Ottawa, Canada, 107-118.
- [57] Vogel J., Chiele, B., 2006. Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39: 897-909.
- [58] 王惠锋, 孙正兴, 王箭. 语义图像检索研究进展. *计算机研究与发展*, 2002, 39(5): 512-523.
- [59] Wang J.Z., Li J., 2002. Learning-based linguistic indexing of pictures with 2-D MHMMs. *Proceedings of the tenth ACM international conference on Multimedia*, 436-445.
- [60] White D.A., Jain R., 1996. Similarity indexing with the SS-tree. In *Proc. 12th Int. Conf. on Data Engineering*, New Orleans, LA.
- [61] Wong Y.M., 2007. Design, implementation, and evaluation of scalable content-based image retrieval techniques, Master Thesis, The Chinese University of Hong Kong.
- [62] 吴洪. 交互式图像检索中的相关反馈技术. 博士学位论文. 中国科学院自动化研究所, 2004.
- [63] 吴洪, 卢汉清, 马颂德. 基于内容图像检索中相关反馈技术的回顾. *计算机学报*, 2005, 28(12): 1969-1979.
- [64] 谢晓霞. CBIR 中相关反馈技术的研究. 硕士学位论文. 北京交通大学, 2006.
- [65] 徐旭. 基于视觉特征的图像检索系统研究. 博士学位论文, 浙江大学, 1999.
- [66] Yu C., 2002. High-dimensional indexing: transformational approaches to high-dimensional range and similarity searches. LNCS. Springer-Verlag Heidelberg.
- [67] Zhang D.S., 2002. Image retrieval based on shape. Phd Thesis, Monash University.
- [68] Zhang C., Chen T., 2002. An active learning framework for content-based information retrieval. *IEEE Trans. On Multimedia*, 4(2): 260-268.
- [69] 章毓晋. 基于内容的视觉信息检索. 北京: 科学出版社, 2003.
- [70] Zhao R., Grosky W.I., 2002a. Negotiating the semantic gap: from feature maps to semantic landscapes, *Pattern Recognition*, 35(3): 593-600.
- [71] Zhao R., Grosky W.I., 2002b. Narrowing the semantic gap-improved text-based web document retrieval using visual features. *IEEE Trans. On Multimedia*, 4(2): 189-200.
- [72] 赵珊. 基于内容的图像检索关键技术研究. 博士学位论文, 西安电子科技大学, 2007.
- [73] 周明全, 耿国华, 韦娜. 基于内容图像检索技术. 北京: 清华大学出版社, 2007.
- [74] Zhu X.Q., Zhang H.J., 2001. New query refinement and semantics integrated image retrieval system with semiautomatic annotation scheme. *J. Electronic Imaging*, 10(4): 850-860.
- [75] 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索. 北京: 清华大学出版社, 2002.

## 第2章 基于内容图像检索关键技术

本章系统地介绍了 CBIR 研究领域的基础知识和若干关键技术，主要内容包括 CBIR 的基本检索原理、图像内容的概念、MPEG-7 标准中的图像特征描述符、CBIR 中常用的低层视觉特征描述方法、图像检索中的匹配技术及图像检索系统中的性能评价准则等问题。

### 2.1 CBIR的基本检索原理

CBIR 建立在计算机视觉和图像理解理论的基础之上，综合了信息检索、图像处理、计算机视觉、机器学习、人工智能、认知心理学、数据库技术等诸多学科的知识。在 CBIR 中，图像内容的描述借助于从图像中自动提取的视觉特征，检索过程是视觉特征间的相似匹配。

假设在基于特征的图像相似性查询中，图像经过特征提取和变换已经转换成为一个  $n$  维向量空间内的点，这样一个图像数据库 DB 就是一个  $n$  维数据空间  $DS$  内的点集，同时，数据库 DB 是完全动态的，允许进行动态的插入和删除。设数据库内对象的个数为  $m$ ，则数据库 DB 可表示为  $\{p_1, \dots, p_i, \dots, p_m\}$ ，其中  $p_i$  ( $p_i = \{p_{i1}, \dots, p_{ij}, \dots, p_{in}\}$ ) 表示数据库中第  $i$  个对象的特征向量， $p_i \in DS$ 。

根据应用领域不同，多维数据库的查询方式也各不相同。对于给定的数据库 DB，基于内容的相似性检索可以划分为三种类型（赵珊 2007，崔江涛 2005，周向东 2003）：

#### 1. 范围查询（Range Query）

给定查询点  $q$ ，查询距离门限  $t$ ，根据距离度量方式  $d$ （距离度量函数），范围查询将查询出所有与点  $q$  距离小于等于  $t$  的点  $p$ 。

$$RangeQuery(q, t) = \{p \in DB \mid d(q, p) \leq t\} \quad (2-1)$$

基于范围的查询方式，其结果集的大小不能事先确定。一个用户在定义了查询距离  $t$  后，并不能确定能得到多少查询结果。这样的话有可能走向两种极端结果，不能得到任何结果，或者得到数据库中的大多数点。在这种情况下，有必要定义一种预先定义结果集大小的查询方式。

#### 2. $k$ 近邻查询（ $k$ Nearest-Neighbor Query）

$k$  近邻查询（又称为  $k$  最近邻查询）是从数据库中选择  $k$  个距离查询点最近的点作为结果集。可以形式化的描述如下：

$$kNN(q, k) = \{p_1, \dots, p_k \in DB \mid \forall p' \in DB \setminus \{p_1, \dots, p_k\}, d(q, p_i) < d(q, p'), 1 \leq i \leq k\} \quad (2-2)$$

在多媒体数据库中的信息检索应用中， $k$  近邻查询是最典型的应用。

#### 3. 限定误差范围内查找（ $\alpha$ -cut）

限定误差范围内查找是指从数据库找出与检索样本的相似度不小于  $\alpha$  的所有图像。

通常范围查询需要比较复杂的检索界面，而且一般要求用户具有关于图像特征的专业

知识与背景,所以这种检索多见于专业图像检索系统(如医学图像检索及诊断系统)。 $k$ -近邻查找和 $\alpha$ -cut 查找是图像检索系统中最常见的检索模式,通常采用简单的 QBE (Query by Example) 检索界面,基于 QBE 的 CBIR 的一般框架见第 1 章图 1.3。

## 2.2 图像内容及检索层次

CBIR 技术的关键是如何描述图像内容,而要描述图像内容,首先必须了解什么是图像内容。本节在分析图像内容及层次的基础上,介绍了 MPEG-7 的常用图像低层特征描述符。

### 2.2.1 图像内容

对于图像数据来说,所谓“内容”具有多个层次上的含义(Hanjalic A. 2001):

#### 1. 感知层 (Perceptive Level)

感知层特征往往指视觉上图像的颜色、纹理、形状、轮廓等特征,这些特征属于图像的低层特征。

#### 2. 认知层 (Cognitive Level)

认知层主要指图像中的主体、对象以及对象间的关系等。认知层特征的提取往往首先要通过图像分割,获取图像中的不同对象,然后提取不同对象的特征以及对象间的关系。

#### 3. 情感层 (Affective Level)

情感层主要是指个人对图像内容的理解,并往往包含个人的情感因素。如印象、情绪、感情等。

黄祥林等将图像内容理解为一个简化了的层次模型,如图 2.1 所示(黄祥林等 2002)。其中第一层是图像的低层物理特征,如颜色、纹理、轮廓和形状等;第二层是逻辑语义特征,反映了图像所描述对象的标识及其空间关系等;第三层是抽象语义特征,是人们对图像内容在认知层次的概括和描述。

上述图像内容层次的划分方法(Hanjalic A. 2001, 黄祥林等 2002, Eakins J.P. et al 1999)实际上具有共同特点,许多学者(Gudivada V.N. et al 1995)把第一种划分方法中的认知层及情感层,以及第二种划分方法中的第二和第三层次概括为“语义层次”内容,而把第一层次和“语义层次”之间的距离称作图像检索的“语义鸿沟”。图像语义的另一个重要特征是它的面向用户的特性,不同知识背景的人有不同的语义需求,并且对于同一幅图像有时会产生截然不同的理解。如何在检索系统中体现这种差别,是图像语义表示的一个重要问题。目前,有关图像的语义特征,主要包括以下 5 种:

- (1) 情感语义。由图像带来的人的感觉,如兴奋,平静,高兴等。
- (2) 行为语义。由图像所表达的行为,如 NBA 球赛,世界杯足球赛等。
- (3) 场景语义。图像所处的场景,如日出,日落,下雨,下雪等。

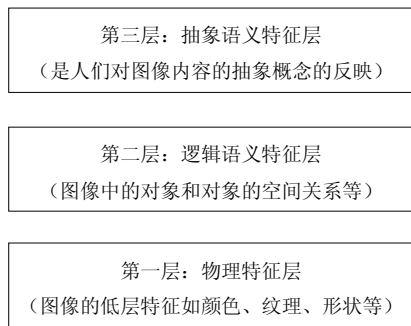


图 2.1 图像特征的层次模型

(4) 空间关系语义。图像中两个或多个对象之间的关系,如 A 在 B 的前面,汽车在房子的后面等。

(5) 对象语义。图像中的对象,如花朵,运动员,房屋,山脉等。

### 2.2.2 图像检索层次

根据图像特征的层次模型,Eakins 等人把用户的检索需求也分为三个层次(Eakins J.P. et al 1999,王惠锋等 2002)。

#### 1. 第一层次

利用图像的颜色、纹理和形状等低层特征及其组合,通过图像的视觉相似性来进行检索。从本质上说,这个层次的检索并没有利用图像的语义信息,目前 CBIR 技术就主要处在这个层次。如“查找一张蓝色占 60% 的图片”,或“检索一张含有椭圆状物体的图片”。

#### 2. 第二层次

根据图像的逻辑特征信息,进行一定的逻辑推理,识别出图像中包含的对象类别。这个层次的检索需求可以是检索一个既定类型的物体,比如“找一张航天飞机的图片”;也可以是检索一个独一无二的人或物,比如“找一张自由女神像的图片”。要让计算机识别某一类的对象,首先必须让计算机“认识”该类对象,即获得对象概念的计算机内部表示,然后找出图像中可能是对象的区域,再来判定对象的类别。对于对象间的空间位置等关系,则是在识别出了对象的基础上来描述它们之间的拓扑关系。

#### 3. 第三层次

根据图像的抽象特征构成检索式,包括物体或场景的描述以及由此推理出来的场景语义、行为语义和情感语义。这个层次的检索需求可以是检索被命名的事件或活动,比如“查找公园玩耍的孩子”,也可以是检索具有情绪特点的图像,比如“查找一张描述高兴的图片”。要回答这一类检索需求,就需要复杂的推理和主观判断,需要抽象的描述图像内容。这类推理和判断往往建立在知识和学习的基础之上,常常要利用心理学和认知科学方面的一系列成果。

这种对用户查询进行层次分类的方法对于描述不同检索技术的能力及其局限性大有帮助。三个层次最主要的差别体现在第一层和第二层之间,即是否真正利用了图像的语义。许多研究者将第二和第三层次的图像检索称为语义图像检索。

Amold 等(Arnold W.M. et al 2000)针对图像检索中存在的“用户从视觉数据中获取的信息与用户自身对视觉数据理解的不一致性”而出现的“语义鸿沟”问题,将图像知识的作用域分成狭义域和广义域。在狭义域中,图像内容变化很小,整个领域的环境也相似,当对象的外观变化较小时,图像的语义描述总体上容易定义,并且大多数情况下是唯一的;而广义域中的图像内容差别较大,它们的语义只能被部分定义。狭义域和广义域的概念可用在刻画使用模式的特征定义和选择及图像检索系统的设计中。在广义域图像中,特征描述和语义注释之间的差距通常较大,而狭义域中的图像、特征和它们的语义注释之间的差距较小,这样,就有可能定义特定的语义模型。同时,他们将图像检索中的知识来源分成以下几个方面:定义图像像素集或图像特征集之间联系的语法等同性和相似性规则;人对等同性和相似性理解的描述规则;在感觉和物体表面特性的差别下描述图像的等同性和差别性的物理规则;描述空间图案等同性和差别性的几何和拓扑规则;基于分类目录的规则;人为的习惯或者与人相关的模式引出的观念等同性和差

别性的规则等方面。

## 2.3 常用的低层视觉特征描述方法

为了度量两幅图像的相似性,首先要从图像中提取代表性的特征,然后采用适当的度量方法比较这些特征的相似度。在理想情况下,应采用与人类的相似性判别相对应的一些特征。由于图像中的物体识别、高级语义的解释等技术的限制,目前尚无法有效地获得这些特征。尽管图像的视觉特征不可能完全描述图像的语义内容,但视觉特征与图像的语义内容之间却存在关联性,只利用图像的低层特征的检索算法,不仅可以检索到一些视觉上相似图像,其中还会包含一些语义相关图像。因此,人们常采用一些稳定的、易于机器提取的低层特征来进行近似性内容描述,如颜色、纹理和形状等,每种视觉特征都具有多种描述方式,每种方式则从不同的角度对视觉特征进行了可计算性描述。

### 2.3.1 颜色特征

颜色特征是在图像检索中应用最为广泛的视觉特征,主要原因在于颜色往往和图像中所包含的物体或场景十分相关。譬如,人们所说的绿色往往和树木草原等相关,谈到蓝色往往和大海蓝天等相关。此外,与其他的视觉特征相比,颜色特征对图像本身的尺寸、方向、视角的依赖性较小,从而具有较强的鲁棒性。因此基于颜色的图像表示方法自然就成为了一种主要的图像索引技术,并得到相当广泛和深入的研究。

孙劲光(孙劲光 2006)将 1995 年以来国内 20 余种核心刊物《计算机学报》、《软件学报》、《计算机研究与发展》、《自动化学报》、《电子学报》、《模式识别与人工智能》、《中国图象图形学报》、《计算机辅助设计与图形学报》、《小型微型计算机系统》、《通信学报》、《CT 理论与应用研究》、《计算机工程与应用》、《计算机应用研究》、《电子测量与仪器学报》、《测绘学报》、《电子与信息学报》、《数据采集与处理》、《信号处理》、《遥感学报》、《中国生物医学工程学报》、《中国体视学与图象分析》发表的 20719 篇学术研究和应用技术文献中属于图像工程研究领域的 563 篇中 58 篇关于基于颜色特征提取与检索技术的文献进行统计分析。并根据学术文章中所提出的学术观点和思想,将上述学术文章进行筛选、归纳和整理,对目前国内图像颜色特征提取方法的研究成果进行分类总结。

鉴于颜色特征在基于内容图像检索中的重要作用,很多基于颜色特征的图像检索算法被提了出来。本章主要从全局颜色特征和空间颜色特征两个方面论述目前国内外主要的颜色特征提取算法。

#### 1. 全局颜色特征

在基于内容的图像检索中,应用最广泛的颜色特征是颜色直方图(Swain M. J. et al 1991)。颜色直方图是先将颜色空间离散化,然后统计图像中每种颜色出现的频率。直方图法容易计算,并且具有图像内容的旋转和平移不变性。采用直方图法作为特征涉及三方面内容,即选择适当的颜色空间,并对此颜色空间进行量化,以及选择相应的相似性判别方法。在基于内容的图像检索中,常用的颜色空间有 RGB、HSV、LUV、YCrCb 及 HMMD(hue-min-max-diff)等。关于颜色空间的详细内容,我们将在第 3 章详细介绍。量化过程中所要考虑的是量化的粒度问题,首先要使颜色空间量化足够精细,保证明显不同的颜色被划分成不同的颜色单元,同时又要使颜色维数尽可能的少,以保证检索的可行

性。由于颜色直方图的相邻单元 (bin) 代表颜色的轻度变化, 对于普通图像而言, 相邻单元的幅值差别不会很大。利用这一性质, MPEG-7 采用 Haar 变换对 HSV 颜色空间直方图进行编码, 用不同的变换尺度和对变换后系数采用的不同量化率获得可缩放的颜色描述子 (SCD), 随着描述比特数的增加, 检索的精确度也会随之增加。Swain 和 Ballard (Swain M. J. et al 1991) 采用直方图相交来度量两个图像内容的相似度, 能部分地排除图像背景的影响, 但并没有考虑量化后不同颜色之间的相关性, 因此对噪声和量化过程中引入的错误具有敏感性。同时, 颜色直方图描述的是图像颜色的统计特性, 丢失了颜色的空间分布信息, 因此, 对于两个颜色直方图相似的图像来说, 如果颜色的空间分布差别很大, 图像的内容会很不相同。

针对上述问题, 近年来许多改进的算法被提了出来。Hafner 等人提出采用二次式距离 (quadratic distance) (Hafner J. et al 1995) 来度量颜色直方图间的相似性, 该方法已被证明比欧氏距离及直方图相交法更为有效, 其原因在于这种距离考虑到了不同颜色之间存在的相似度。文献 (Li J. et al 2000, 卢汉清 2000) 采用彩色图像的主要颜色来构造颜色直方图, 由于忽略了那些数值较小的颜色区间, 改进后的颜色直方图对图像噪声的敏感程度降低了, 从而使检索效果更好。文献 (Stricker M. et al 1995, 刘忠伟 1998) 利用了直方图中各个颜色的位置关系, 提出采用累加直方图方法和局部累加直方图来提高采用直方图特征进行图像检索的效率。由于直方图中位置相近的颜色要比远离的颜色的相似度高, 因此这种方法即使在粗糙的量化情况下也能具有较好的稳健性。Guoping Qiu 等人 (Qiu G.P. et al 2003) 根据人的视觉特性与频域中频率的关系, 提出采用高通滤波器、低通滤波器以及带通滤波器将图像的亮度分量转化为不同的频率分层, 进一步将彩色图像转化为不同的频率分层, 然后对于每一频率分层采用颜色直方图方法来进行索引, 从而有效的提高图像的检索效果。

另一种非常简单且有效的颜色特征描述方法可以使用 Stricker 和 Orengo 所提出的颜色矩 (color moments) (Stricker M. et al 1995)。这种方法的数学基础在于图像中任何颜色分布均可以用它的矩来表示, 并且极低的特征维数使其具有很强的竞争力; 根据颜色直方图特性和信息论中信息熵的概念, John (John ZM. 2000) 提出采用图像颜色的信息熵来表示图像的颜色特征, 将图像的颜色直方图由多维降低到一维。但是, 对于颜色熵和颜色矩来说, 对于视觉不同的直方图, 可能具有相同的值, 为了解决该问题, 我们提出了相应的改进办法 (孙君顶等 2005, Sun J.D. 2006) (详细方法见的第 3 章)。

## 2. 空间颜色特征

上述颜色直方图、颜色矩和颜色熵等描述符所描述的颜色特征是图像的全局颜色特征, 不包括图像颜色的空间分布特征, 因此仅仅利用这些特征进行图像检索时, 极易造成误检现象。基于此, 许多用于提取图像颜色空间分布特征的方法被提了出来。

Pass 等人 (Pass G. et al 1996) 提出了颜色聚合向量的方法, 此方法将颜色直方图中的每个统计单元分为连贯和非连贯两类。一个像素如属于一块连续一致的颜色区域则认为其连贯, 否则为不连贯。Huang 等通过在直方图中结合颜色的位置相关性提出了颜色相关图 (color correlograms) (Huang J. 1998), 此方法统计某种颜色的像素在与其相距某固定距离下出现各种特定颜色像素的概率。由于该方法运算复杂度过高, 一般只统计少数几个局部距离下的同一颜色出现概率的自相关图。文献 (Rao A. et al 1999, Cinque L. et al 1999, Lim S. et al 2003) 提出了更一般的颜色几何分布直方图 (环形颜色直方图、角度颜色分布

直方图等), 这些方法预先指定一组像素分布的几何形状, 然后在图像中统计符合这种几何形状分布的一定颜色出现的频率。在上述方法的基础上, 我们结合信息熵理论, 提出了基于颜色分布熵的颜色特征描述算法 (Sun J.D. et al 2006b)。为了增加局部颜色分布信息, MPEG-7 在 HMMD 颜色空间内构造了一个颜色结构直方图 (CSD)。在每个像素位置处, 与传统的直方图统计方法不同, CSD 统计一个像素滑动窗口内出现的颜色种类, 以此来区别那些具有相似的颜色直方图, 但具有不同颜色空间分布的图像。

结合空间结构关系的另一种方法是将图像划分为预先指定的多个区域, 通过比较各个区域的直方图来提高空间判别能力。局部颜色特征目前从划分局部区域的角度来说, 常用的划分方法包括: 基于固定块的图像分割、基于手工的区域分割、采用交互半自动的区域分割以及一些自动的颜色分割方法。同时, 局部区域颜色信息主要采用平均颜色、主颜色、颜色直方图、颜色矩和二进制颜色集等来表示 (李向阳等 2001, Hsu W. et al 1995, Dimai A. 1997, Stricker M. et al 1997, Sciascio E.D. et al 1999, Sebe N. et al 1999, Stehling R.O. et al 2000, 何清法等 2001, Yoo H.W. et al 2002)。如, Stricker 和 Dimai (Stricker M. et al 1997) 认为图像中最有意义的区域是位于图像的中心, 因而将图像划分为 5 个模糊的区域, 其中中心区域为一个椭圆形, 同时将椭圆外的图像进行一定的模糊化, 以确保图像中的颜色信息对于图像小的平移和旋转影响不敏感, 然后采用每一区域的颜色矩作为图像的索引特征。Sebe (Sebe N. et al 1999) 采用多分辨率的图像空间划分方法, 使用不同分辨率的网格将图像划分为不同等级。在进行图像划分时, 各分块间采用可重叠的方式, 然后提取每一分块的主颜色或颜色矩作为该分块的颜色特征。由于使用不同尺寸的矩形, 该方法能够处理图像中不同尺寸的物体和区域, 而且可互相重叠的分块划分方法使其在给定尺度下图像中的所有物体都能被一个区域包含, 因此提高了检索的精度。

### 2.3.2 纹理特征

纹理是图像中一个重要而又难以描述的特征, 至今没有公认的定义, 但却受到了广泛的研究和重视。一种定义是把图像中局部不规则, 而在整体上表现出某种规律的特征称之为纹理。纹理描述的是事物表面上的一种视觉特征, 它是由非单一颜色或明暗度所呈现出的具有同质属性的视觉模式, 例如云彩、树木、砖、织物等都有各自的纹理特征。纹理蕴含着丰富的视觉信息, 尽管还没有一种确切的定义, 但纹理能很容易地被人类所感知, 并且在机器视觉领域得到了广泛地研究, 而且人们发现, 纹理特征具有一种不依赖于颜色或亮度的反映图像中同质现象的视觉特征, 可以从微观上区分图像中不同的物体。

由于纹理特征对模式识别和计算机视觉等领域的重要意义, 在图像检索中, 纹理特征也发挥了很大的作用, 近年来取得了越来越多的研究成果。常用的纹理描述方法有: 统计法、频谱法、结构法和模型法四种 (Pentland A. et al 1994, Liu F. et al 1996, Kuan J.P. et al 1999, Sebe N. et al 2001, 汪祖媛等 2001, 黄元元等 2003, 吴冬升等 2003, 赵珊等 2006, 赵珊等 2007)。

#### 1. 统计法

统计法分析纹理的主要思想, 是通过图像中灰度级分布的随机属性来描述纹理特征。最简单的统计法是借助灰度直方图的矩来描述纹理 (Lowitz G.E. 1983), 但这种方法没有利用像素相对位置的空间信息, 为了利用这些信息, Haralick 等人提出了利用共生矩阵来描述纹理特征 (Haralick R.M. et al 1973), 该方法研究了纹理的空间灰度级相关性, 构造

了一个基于像素间方向和距离的共生矩阵, 然后从灰度共生矩阵中提取有意义的统计特性作为纹理特征的表达, 所提取的统计特征主要包括能量、熵、相关性、惯量及纹理能量等。基于人类对纹理的视觉感知和认知心理学的研究, Tamura 等人 (Tamura H. et al 1978) 提出了相应的纹理特征表达方式, Tamura 纹理特征的六个分量分别对应于心理学角度上纹理特征的六种属性: 对比度 (contrast)、粗糙度 (coarseness)、方向性 (directionality)、线像度 (line likeness)、规整度 (regularity) 和粗略度 (roughness)。粗糙与精细相对应, 体现了纹理中具有重复性的纹理单元尺寸的一种近似性度量; 对比度是体现一个模式的鲜明程度的; 方向性描述了纹理模式中是否存在明显的方向性信息; 线象性与斑块性相对应, 体现了纹理元素的形状信息; 规则性是指纹理元素排列的规律性; 粗略与光滑相对应, 体现的是一种触觉特性。鉴于这些特征具有一定的视觉感知意义, 在 QBIC (Niblack W. et al 1993) 和 MRAS (Huang T.S. et al 1996) 系统中采用了改进后的 Tamura 特征。除上述纹理描述外, 纹理谱 (He D.C. et al 1991) 也是目前广泛应用的一种纹理特征描述方法, 其主要思想是利用局部纹理模式作为纹理单元来进行纹理分析, 通过刻画图像像素点邻域内灰度的变化来描述图像的纹理结构特征。在此基础上, Ojala 等人 (Ojala T. et al 1996) 提出了 LBP (Local Binary Pattern) 方法, 该算子在纹理分类上取得了很好的效果, 近年来, 广泛应用于医学图像处理及人脸识别等图像处理领域。纹理谱方法与以往的方法相比概念清楚, 所需的计算量较小, 近年来越来越受到重视。有关 LBP 方法的应用情况还可参考文献 (Ojala T. et al 1996, 1998, 1999, 2000, 2002) 等, 更多的内容可查阅网站 ([http://www.ee.oulu.fi/mvg/page/lbp\\_bibliography](http://www.ee.oulu.fi/mvg/page/lbp_bibliography))。文献 (万华林等 2003, 施智平等 2004, 木拉提·哈密提等 2006) 在 LBP 的基础上, 进行了相关改进。在 LBP 的基础上, Marko 等人提出利用 LBP 来描述感兴趣区域的特征 (Marko H. et al 2009)。

## 2. 频谱法

频谱法主要借助于频率特性来描述纹理特征。常用的频谱法主要包括傅里叶功率谱法 (章毓晋 2003)、Gabor 变换 (Manjunath B.S. et al 1996)、塔式小波变换 (Pyramid Wavelet Transform, PWT) (Michael D. 2001)、树式小波变换 (Tree Wavelet Transform, TWT) (汪祖媛等 2001) 等。近年来, 小波理论成为多尺度 (多分辨率) 图像分析方法的一种正式和统一的数学框架, 滤波器组和小波分解描述方法一般采用图像的各个子带的输出能量作为纹理特征 (Randen T. et al 1999)。

## 3. 结构法

结构法分析纹理的基本思想是, 假定纹理模式由纹理基元以一定的、有规律的形式重复排列组合而成, 特征提取就变为确定这些基元并定量的分析它们的排列规则 (Young D.C. et al 2003)。由于纹理基元描述了局部纹理特征, 因此对整幅图像中不同纹理基元的分布统计可获得图像的全面纹理信息 (章毓晋 2001)。纹理基元可能是明确的、直观的, 也可能是不明确的、需要人为的根据情况来设定的。无论如何确定纹理基元, 对于存在纹理基元之间的结构关系, 可以有不同的分析途径 (Zhou X.S. et al 2001)。最简单的方法是分析纹理基元之间存在的相位、距离、尺度等统计特性, 也可以考虑用复杂的方法分析, 如句法分析等。具体方法的选择依赖于纹理分析任务的要求。结构法是最符合纹理本质的纹理分析方法, 但由于目前图像分割技术尚无法准确提取纹理基元, 因此该类方法比较适合规则的人工纹理。



#### 4. 模型法

模型法主要有随机场方法和分形法两种。这些模型的共同特点是通过少量的参数表征纹理。常见的随机场模型有Gauss-Markov模型 (Rellier G. et al 2004, Paget R. 2004)、Gibbs模型 (Li C.T. 2003, Gimelfarb G. 1999), 随机场方法的缺点是参数难估计, 计算量大, 且自然纹理很难用单一的模型表达。分形法则是利用分形维数进行纹理粗糙度的度量 (Minghong P. et al 2005), 许多自然图像的粗糙度具有一些统计性质, 在不同的尺度上具有自相似性, 分形对于这些性质建模非常有用。但分形维数的计算一般是采用一个理论模型, 而实际的纹理图像并不完全符合理论模型。另外, 不同的纹理图像可能具有相似的分形维数, 使进一步的处理复杂化 (Hou Z. et al 2003)。关于模型法的研究, Markov随机场 (Markov Random Field, MRF) 模型取得了很大的成功, 其中自回归纹理模型 (Simultaneous Auto-Regressive, SAR) 就是MRF模型的一种应用实例 (Khotanzad, A. et al 2003)。

#### 2.3.3 形状特征

形状是图像的重要可视化内容, 是人类视觉系统进行物体识别时所需要的关键信息之一。它不随周围环境, 如亮度等的变化而变化, 是物体的稳定信息。人们对一幅图像的理解很大程度上有赖于对图像中目标形状的区别和感知。然而, 与颜色、纹理特征不同, 物体形状的自动获取及匹配比较困难, 目前的技术尚无法达到精确的图像分割, 图像分割在理论上仍缺乏简单易行又准确可靠的广义的方法, 因此, 基于形状的检索一般仅限于容易识别的物体 (Faloutsos C. et al 1994)。就现在的检索应用而言, 在一般图像库中即使采用了形状特征, 也很少利用全自动分割技术。另一方面, 形状通常包含了非重复性空间变化的细节信息, 这无疑给形状特征的描述增加了负担, 至今还没有找到与人对形状的感觉完全一致的对形状的确切的数学定义。而且在形状相似性判别方面, 也缺乏严格的数学定义来解释人类对形状相似性判别的感知性。从不同视角获取的图像中的目标形状, 可能会有很大差别, 为了准确进行形状匹配, 还需要解决受噪声污染、平移、尺度、旋转变换以及闭塞等的不变性的问题。以上原因都使形状检索成为图像检索领域中一个具有挑战性的课题。而对于一些专门用途的图像而言, 如商标、手工工具、海洋鱼类、已分割好的物体形状或物体轮廓图等, 这些图像的背景一致、前背景对比明显, 图像分割已不再是阻碍形状特征应用的一个难题。因此, 与其他方法相比, 在针对此类图像的检索应用中, 形状已成为一种有效和最具潜力的检索手段。

在二维图像空间中, 形状通常被认为是一条封闭的轮廓曲线所包围的区域。因此, 形状特征包括轮廓和区域的描述, 前者只用到形状的外边界, 而后者则关系到整个形状区域。从方法上来讲, 图像的形状视觉特征提取算法可以分为基于轮廓和基于区域两种。

##### 1. 基于区域的描述方法

基于区域的形状描述符是利用区域内的所有像素集合来获得描述目标轮廓所包围的区域性质的参数。常用的方法有几何不变矩 (Hu M.K. 1962), Zernike 矩 (Yang L. et al 1994), 广义傅里叶描述 (Zhang D.S. 2002) (Generic Fourier Descriptor, GFD), 小波重要系数法 (Jia Li et al 2000) 和网格法 (Lu G.J. et al 1998, Lu G.J. et al 1999)。Hu (Hu M.K. 1962) 基于形状不变矩 (Moment invariants) 提出了一系列分别具有变换、旋转和缩放无关性的 7 个矩。使用 7 个矩的计算结果构成七维向量来表示形状, 该方法具有计算快速和存储开销少的特点。在此基础上, 文献 (Adoram M. et al 1999) 采用梯度向量流主动

轮廓线寻找形状边缘,由此计算出几何不变矩作为形状特征向量,并用向量间的欧氏距离作为相似性测度。Teague (Teague M.R. 1980) 采用了更一般形式的 Legendre、Zernike, 伪 Zernike 等正交多项式作为矩变换核代替传统的矩变换核,在这种情况下矩产生了最小的信息冗余。Yang (Yang L. et al 1994) 根据正交多项式的原理提出用正交矩来恢复图像,并引进了 Zernike 矩。其思想和傅里叶变换思想类似,将信号展开成一系列的正交基,但 Zernike 矩无需知道边界信息,更适合于描述复杂形状。与基于几何不变矩的方法相比,Zernike 矩可以最高计算到 36 矩阶,具有极强的形状分辨能力,但高阶矩的计算复杂,目前常使用的为 Zernike 矩的前 10 阶矩。由于 Zernike 矩的计算需要首先对图像进行归一化处理,而在归一化处理的过程中会对图像的像素进行重新取样和映射,因此引入了新的误差。Kamila (Kamila N.K. et al 2005) 提出使用改进的 Zernike 矩进行特征提取,即(伪 Zernike 矩)解决重取样的误差问题。广义傅里叶变换是将 2D 傅里叶变换应用于使用极坐标表示的图像。该方法易于计算,是纯频率域特征,在形状区域的径向和角度可以做多分辨率分析。Zhang (Zhang D.S. 2002) 在其博士论文中提出了广义傅里叶描述符(GFD)的算法,该算法对图像的像素在极坐标上进行 2D 傅里叶变换,并使用变换后的傅里叶系数为特征向量,此方法对图像的平移和缩放具有较好的鲁棒性。Zhang 将该方法与目前 MPEG-7 标准中推荐的 Zernike (Yang L. et al 1994) 算法相比,通过实验得到该方法优于 Zernike 算法的结论。Lu 等人 (Lu G.J. et al 1998) 使用一个最小外包矩形覆盖整个形状,并通过网格将外包矩形划分成固定的单元格;依据每一个单元格与形状的覆盖情况,超出阈值则设置为 1,否则设置为 0;通过自上而下,自左向右的方式顺序扫描作为形状的特征描述,然后利用异或(XOR)操作进行图像的匹配。该方法计算简单、容易实现,但是对于形状的旋转变换敏感。在网格描述符的基础上,我们提出了基于信息熵的图像检索算法(孙君顶等 2004b)。

除此之外,MPEG-7 的 ART 描述子在极坐标的单位圆面上采用角径向变换(Angular Radial Transformation)来获得矩不变量,量化 ART 基函数的系数作为描述特征,并用于匹配。在 QBIC 系统中采用了全局特征用于形状的描述,如面积、圆度、偏心率、离心率、主轴方向、代数矩等,并采用加权欧氏距离作为相似度测度。除上述描述符外,常用的基于区域的描述符还有区域的面积、欧拉数、离散度、偏心率、区域骨架、几何不变矩、旋转矩、复数矩、角半径变换等方法等。

## 2. 基于轮廓的描述方法

基于轮廓的形状描述符是对包围目标区域的轮廓的描述。基于边缘的图像特征一般可采用谱描述子、尺度空间滤波、多边形近似等一些方法。在闭合轮廓线上的任意一点位置可以看做为一个周期函数,则此轮廓线可用 Fourier 级数展开,低频信息对应轮廓的概要特征,高频信息对应轮廓的细节特征。进行 Fourier 变换先要对轮廓线采样,再进行离散 Fourier 变换,变换后的系数通常叫做 Fourier 描绘子(Zhang D. S. 2002)。Fourier 描绘子很容易进行归一化,以获得一组具有旋转、平移和尺度不变性的特征向量,再用特征向量之间的欧氏距离作为相似度测度。为了更好地获得局部形状特征,一些研究者采用了短时 Fourier 描述子和小波描述子(杨翔英等 1999)。傅里叶描述子方法常用的还有基于极半径函数法(Persoon et al 1997),基于曲线转角函数法(Zahnand et al 1972, Granlund 1972)等。Mokhtarian 等人引入曲率尺度空间描述符(Curvature Scale Space Descriptor, CSSD)来描述轮廓特征(Mokhtarian F. et al 1992),并成功应用于形状特征的描述及检索中

(Mokhtarian F. et al 1996a, Mokhtarian F. et al 1996b)。目前, 该描述符成为MPEG-7 标准中轮廓描述的一种方法。该方法的基本思想是基于人们在认知物体时, 倾向于将物体的形状分解成凹和凸的部分来比较这一特性 (Manjunath B.S. et al 2002)。CSSD是在多尺度下根据变形点将轮廓“分解”成凹和凸部分, 这些变形点定义为轮廓曲线上曲率为零的点。

除上述轮廓描述方法外, 常用方法还有Freeman链码 (Iivarinen J. et al 1996) 及基于链码的改进方法 (王小玲等 2006, Sun J.D. et al 2006a, Sun J.D. et al 2007, 孙君顶 2008)、hausdorff distance (Chetverikov D. et al 1999, Meier T. 1998, Rucklidge W.J. 1997, Scassellati B. et al 1994)、shape contexts (Belongie S. et al 2001)、BAS (Beam angle statistics) 描述符 (Arica N. et al 2003)、TAR (Triangel-Area Representation) 描述符 (Alajlan N. et al 2007)、边界矩 (Chen, C.C. 1993, Gupta et al 1987, Sun J.D. et al 2008) 等。简单的边界几何形状不变量, 如周长、长轴、短轴、主轴方向等都可以用来描述形状的某一个特征, 且具有旋转不变等特性。但是它们的提取依赖于图像的分割, 而且噪声也会对最终的计算结果造成巨大的影响。

### 2.3.4 MPEG-7 中的图像特征描述符

面对如何有效地识别、过滤、浏览和检索视听材料的迫切要求, MPEG 于 1996 年开始从事一项新的工作, 目的是为多媒体内容的描述提供一个标准。MPEG 的这个新成员被正式命名为“多媒体内容描述接口 (Multimedia Content Description Interface)”, 简称为 MPEG-7。其目标是产生一个描述多媒体内容的标准, 支持对多媒体信息在不同程度层面上的解释和理解, 从而使其可以根据用户的需要进行传递和存取。MPEG-7 并不面向某种具体的应用, 相反, MPEG-7 标准将支持尽可能广泛的应用领域。MPEG-7 标准中确定了一个标准描述子 (Descriptors) 集, 用来描述各种类型的多媒体信息。描述子是特征的描述, 它定义了特征描述的句法和语义。一个特征, 如颜色、纹理或形状, 可能多个描述子对其不同相关方面进行描述。如图 2.2 所示,

MPEG-7 注重的是提供视听信息内容的描述方案, 并不包括针对不同应用的特征提取方法和搜索引擎, 这使得 MPEG-7 标准一方面可以被广泛的应用, 不局限于某些与特殊应用密切相关的特征提取



图 2.2 MPEG-7 的范围

算法和搜索引擎, 也不依赖于被描述内容的编码和存储方式; 另一方面又可以引入竞争机制, 使人们能够针对不同应用领域产生出更多更好的特征提取算法和搜索引擎。有关 MPEG-7 的更详细介绍可参阅 (<http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>) 及 ISO/IEC 颁发的有关 MPEG-7 的文档。

MPEG-7 标准中包含几个重要概念: 描述符 (Descriptors, Ds)、描述方案 (Description Schemes, DS) 及描述定义语言 (Description Definition Language, DDL)。其中, 描述符表示多媒体特征信息的语法和语义属性。一个特征可以用多个描述符来表示, 比如, 平均色 (average color)、主色 (dominant color) 和颜色直方图 (Color Histogram) 等都是颜色特征的描述符。描述方案指定了对象或者特征的结构和关系, 一般情况下, 描述方案是解决图像分类和组织问题, 或以特定的索引结构描述图像内容的子系统。描述定义语言允许创建新的描述方案和描述符, DDL 允许扩展和修改现存的描述方案。目前, MPEG-7 采用带有 MPEG-7 特定扩展的 XML (可扩展置标语言) 语言作为描述定义语言。

在 MPEG-7 标准中考虑了五类基本的视觉特征, 对应的使用了五类描述符: 颜色描述符、纹理描述符、形状描述符、运动描述符和位置描述符。其中, 前三类描述符都是针对单幅静止图像的描述。由于 MPEG-7 的目的是建立对图像信息完备而又一致的描述, 因此, MPEG-7 中提出的颜色描述符、纹理描述符和形状描述符等单类描述符必须尽可能的完备而又一致。也就是说, 使用这些描述符能尽可能完备地表达图像中的视觉信息, 并且更重要的是这些描述符适用于所有图像。

下面主要介绍一下 MPEG-7 中有关图像内容描述三类描述符: 颜色描述符、形状描述符及纹理描述符 (Chang, S.F. et al 2001, Ebrahimi T. et al 2001, Koenen, R. et al 2000, Manjunath B.S. et al 2002, Bober, M. 2001, Sung M.K. et al 2007, Manjunath B.S. et al 2001, 王郑耀 2006, 帅勤 2006)。

### 1. 颜色描述符

MPEG-7 中的颜色描述符有颜色空间描述符 (Color Space Descriptor)、颜色量化描述符 (Color Quantization Descriptor)、主颜色描述符 (Dominant Color Descriptor)、可伸缩颜色描述符 (Scalable Color Descriptor)、颜色布局描述符 (Color Layout Descriptor)、颜色结构描述符 (Color-Structure Descriptor) 及帧图/图组颜色描述符 (Group of Frames/Group of Pictures Color Descriptor) 等。其中, 颜色空间和颜色量化是两个辅助性的颜色特征描述符, 它们不能独立使用, 只能配合其他四个描述符使用。

颜色空间描述符描述了 MPEG-7 颜色描述符的颜色空间, 包括 RGB、YCbCr、HMMD、HSV 以及各种颜色系统与 RGB 的线性变换矩阵。

颜色量化描述符描述了颜色空间的均匀量化, 量化产生的维 (Bin) 的数目是可配置的, 这样使得各种应用具有更大的灵活性。该描述符往往需要和主颜色等描述符配合使用。

主颜色描述符最适用于表示局部 (对象或图像区域) 特征, 几种颜色就足以表达我们感兴趣区域的颜色信息。当然, 它也可以用于整个图像, 例如旗帜图像或彩色商标图像。颜色量化用于提取每个区域/图像的少数代表颜色, 并相应的计算出区域中每种量化颜色所占的百分比, 同时该描述符还定义了整个描述符的空间相关性, 用于相似性检索。

可伸缩颜色描述符定义了 HSV 空间的颜色直方图, 然后用 Harr 变换编码。根据 Bin 的数目和 Bit 表示的精度, 它的二进制表达在 Bin 的数量和 Bit 表达精度上都是可伸缩的。这个描述符主要用于图像与图像的匹配和基于颜色特征的检索, 检索的精度随着描述中使用的比特数目的增加而增加。

颜色布局描述符描述了整幅图像或者图像的部分区域的颜色空间分布状况。该描述符以一种紧凑的形式, 有效的表达了颜色的空间分布。这种紧凑性以很小的计算代价带来高速的浏览和检索。它提供图像与图像的匹配和超高速的片断与片断的匹配, 这些匹配要求大量相似性计算的重复。由于该描述符表达了颜色特征的布局信息, 因此它可以提供相当友好的用户接口, 例如使用其他颜色描述符中均不支持的手绘草图查询。

颜色结构描述符是一个颜色特征描述符, 它既包括颜色内容信息 (类似于颜色直方图), 又包括内容的结构信息。它的主要功能是图像与图像的匹配, 主要用于静态图像检索。在这里一幅图像可能由一个单一矩形或者任意形状, 也可能是非连通的区域组成。提取方法是通过考虑一个  $8 \times 8$  像素的结构化元素中的所有颜色, 将颜色结构信息加入到该描述符中, 而不是单独考虑每个像素。

帧图/图组颜色描述符用于静态图像的可伸缩颜色描述符扩展到对视频片断或静态图

像集合的颜色描述。在 Haar 变换之前,用附加的两个比特定义如何计算颜色直方图:是均值、中值还是相交。

## 2. 形状描述符

MPEG-7 中定义的形状描述符有区域形状描述符 (Region Shape Descriptor)、轮廓形状描述符 (Contour Shape Descriptor) 及 3D 形状描述符 (Shape 3D) 三种。

物体的形状可能是一个独立的整体,也可能是由几个区域构成的。由于使用了图像中构成物体形状的所有像素信息来描述物体,区域形状描述子不仅仅可以用来刻画出个别独立的连通区域,也包括由一些不连通区域所构成的复杂图形,如图 2.3 所示。

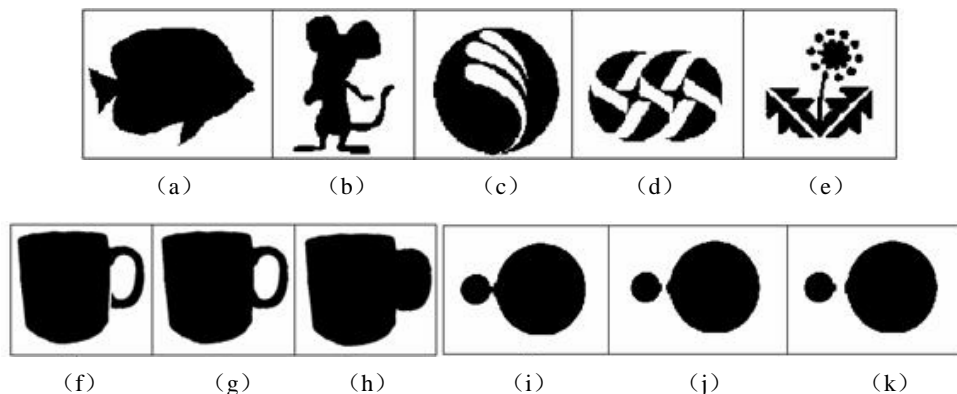


图 2.3 各类形状示例图

区域形状描述符的表达式是由一系列 ART (Angular Radial Transform) 系数构成的。ART 定义了一组二维的复值正交基函数,将二维区域投射到这些基函数上,得到的系数归一化后就可以描述区域的形状并用于匹配。区域形状描述符也是一种非常紧凑、有效的描述方式,并具备分割噪声的功能。

轮廓形状描述符是利用轮廓的多尺度曲率空间 (Curvature Scale Space, CSS) (Farzin Mokhtarian et al 1992) 来描述封闭的轮廓。具体地说,对封闭轮廓线进行多尺度变换,得到多个尺度的轮廓线,随着尺度的变换,轮廓线越来越平滑。多尺度变换可以使用高斯函数对曲线上的点进行平滑来完成。然后,在每一个尺度计算曲线的曲率,得到曲率变化的零交叉点,使用这些点就可以描述轮廓的形状。

MPEG-7 三维形状描述符可用于相对自然的或虚拟的三维目标。在描述三维物体的形状特征时,首先建立物体的三维网格 (3D Mesh) 模型,然后在物体表面的局部区域计算出该处的形状指数 (Shape Index)。该描述符基于形状频谱 (Shape Spectrum) 的概念,事实上就是基于形状指数的直方图。三维形状描述符常常用于三维物体的匹配。

## 3. 纹理描述符

MPEG-7 中的纹理描述符包括同质纹理描述符 (Homogenous Texture Descriptors)、纹理浏览描述符 (Texture Browsing Descriptors) 和边缘直方图描述符 (Edge Histogram Descriptors) 三种。同质纹理描述符在纹理具有一致性的区域统计纹理的空间频率;纹理浏览描述符从纹理的方向性、规则性和粗糙程度三方面进行描述;边缘直方图描述符在纹理不具有一致性的区域描述了边缘的空间分布。

同质纹理作为一个重要的视觉基本特征,它通过在频域计算能量 (energy) 和能量方

差 (energy deviation) 来提供对纹理的量化描述。同质纹理描述子采用了 5 个尺度和 6 个方向的 30 个 Gabor 滤波器对纹理图像进行多分辨率分解, 将频域内滤波器组输出能量的均值和标准差作为纹理特征。它描述了图像中与人类感知一致的规则性、方向性和粗糙度等纹理特性, 最适合对具有同质特性纹理进行一定的描述, 可用于纹理图像数据库中图像之间的相似性匹配, 因此主要用于大量相似图案的搜索和浏览。一幅图像可看做由同质纹理以马赛克形式拼接而成的, 所以与这些区域关联的纹理特征可以作为索引来检索图像。例如, 用户浏览一个航空图像数据库, 可能想识别图像集合中的停车场, 当从远处观察时, 汽车规则 (以相等间隔) 停放的停车场就是一个极好的同质纹理图案的例子。同样的, 从空中或是卫星拍摄的农田和植被图像也是同质纹理的示例。该描述符可以支持诸如“找出一个看起来与这个区域相似的植被”的查询。为了支持这样的图像检索, 需要对纹理进行有效的表示。

纹理浏览描述符从类似于人类感知的角度对纹理的方向性 (Directionality)、规则性 (Regularity) 和粗糙程度 (Coarseness) 进行了描述, 适用于图像的浏览和根据纹理粗糙程度进行的分类。由于一个纹理可能不只包含一个主要方向和相应的尺度, 因此允许最多指定两个不同的方向和粗糙度值。该描述符的计算方法和同质纹理描述符类似, 首先使用一组带有方向和尺度参数的 Gabor 滤波器进行滤波, 然后通过分析滤波结果, 找到纹理主要的方向; 接着分析滤波后的图像沿着这两个 (第二个主方向是可选的) 主方向投影, 以此决定纹理的方向性、粗糙度和规则性。同质纹理描述符和纹理浏览描述符提供了表示相似纹理区域的多尺度方法。这是一个非常紧凑的描述子, 适于浏览应用和纹理过滤, 与同质纹理描述子联合使用可以实现快速准确的图像检索。

边缘直方图描述了边缘的空间分布信息。首先将图像划分成 16 个互不重叠的矩形区域, 对每个图像区域分别按水平、垂直、45°、135° 四个方向和一个无方向性边缘五类信息 (如图 2.4 所示) 进行直方图统计。此描述子具有尺度不变性, 支持纹理旋转和旋转不变匹配, 适用于非一致纹理图像, 如普通图像的检索。

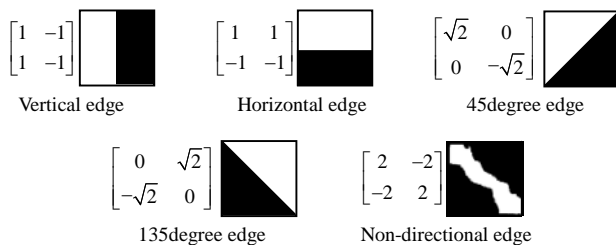


图 2.4 边缘类型定义

## 2.4 特征匹配技术

图像检索的匹配策略大致可以分为两种, 一种是完全匹配, 另外一种是相似性匹配。当两幅图像的特征完全相同时, 图像匹配成功, 称为完全匹配。当两幅图像的特征间的距离小于某一个阈值时, 图像匹配成功, 称为相似匹配。传统数据库中的完全匹配并不是图像数据库检索的主流 (往往用于基于图像元数据的检索), 在基于内容的图像检索中, 占

主导地位的是建立在图像低层视觉特征对比基础上的相似性检索。在提取图像特征后,可采用相应的相似度量策略来进行特征匹配,也就是通过确定检索图像同数据库目标图像特征向量间的距离来确定待检索图像同数据库中目标图像间的相似性。

在基于内容的相似性检索的三种类型(见 2.1 节)中, $K$ -最近邻查找和限定误差范围内查找( $\alpha$ -cut)都与相似度量函数的定义有密切的关系。显然,一个合适的相似度量方法对图像检索结果影响很大。相似度量方法的好坏会影响到图像检索的性能,相似度度的计算复杂度会影响到用户对图像检索的响应时间。理想的相似度量方法应该满足人的视觉特征,也就是说视觉上相似的图像间应具有较小的距离,而视觉上不相似的图像间应具有较大的距离(孙君顶 2005)。

### 2.4.1 计量定理

设  $A, B, C$  为任意的  $n$  维特征向量,通常情况下,距离度量函数  $d$  应受以下四条公理的限制(章毓晋 2003, Zhang D. S. 2002, Santini S. et al 1996):

自相似公理

$$d(A, A) = d(B, B) = 0 \quad (2-3)$$

最小公理

$$d(A, B) \geq d(A, A) = 0 \quad (2-4)$$

对称公理

$$d(A, B) = d(B, A) \quad (2-5)$$

三角不等公理

$$d(A, C) \leq d(A, B) + d(B, C) \quad (2-6)$$

在实际应用中,所采用的相似度比较函数并非严格满足上述距离度量的四条公理,它们往往只是满足上述公理的某个或某几个。

### 2.4.2 常用的匹配算法

目前,图像检索中用到的特征匹配算法很多,常用的有以下几种(孙君顶 2005)。

(1) Minkowsky 距离

Minkowsky 距离是基于  $L_p$  范数定义的,即

$$L_p(A, B) = \left[ \sum_{i=1}^n |a_i - b_i|^p \right]^{\frac{1}{p}} \quad (2-7)$$

如果  $p=1$ ,  $L_1(A, B)$  称为城区(city-block)距离,即

$$L_1(A, B) = \sum_{i=1}^n |a_i - b_i| \quad (2-8)$$

如果  $p=2$ ,  $L_2(A, B)$  称为欧氏距离(Euclidean distance),即

$$L_2(A, B) = \left( \sum_{i=1}^n (a_i - b_i)^2 \right)^{\frac{1}{2}} \quad (2-9)$$

如果  $p \rightarrow \infty$ ,  $L_\infty(A, B)$  称为 Chebychv 距离,即

$$L_\infty(A, B) = \max \left( \sum_{i=1}^{\infty} |a_i - b_i| \right) \quad (2-10)$$

## (2) 直方图相交法

直方图相交法 (histogram intersection) 是由 Swain 等人于 1991 年首次提出的, 直方图相交法计算简单快速, 并且能较好地抑制背景的影响, 其数学描述为 (Swain M.J. et al 1991)

$$d(\mathbf{A}, \mathbf{B}) = 1 - \sum_{i=1}^n \min(a_i, b_i) \quad (2-11)$$

式 (2-11) 可以进一步进行归一化处理 (庄越挺等 2002) 为

$$d(\mathbf{A}, \mathbf{B}) = 1 - \frac{\sum_{i=1}^n \min(a_i, b_i)}{\min(\sum_{i=1}^n a_i, \sum_{i=1}^n b_i)} \quad (2-12)$$

## (3) 二次式距离

对于基于颜色直方图的图像检索来说, 二次式 (quadratic distance) 距离 (Hafner J. et al 1995) 已被证明比欧氏距离及直方图相交法更为有效, 其原因在于这种距离考虑到了不同颜色之间存在的相似度。二次式距离可以表示为

$$d_{\text{quad}}(\mathbf{A}, \mathbf{B}) = (\mathbf{A} - \mathbf{B})^T \mathbf{M} (\mathbf{A} - \mathbf{B}) \quad (2-13)$$

式中,  $\mathbf{M} = [m_{ij}]$ ,  $m_{ij}$  表示直方图中下标为  $i$  和  $j$  的两种颜色之间的相似度。这种方法通过引入颜色相似性矩阵  $\mathbf{M}$ , 使其能够考虑到相似但不相同的颜色间的相似性因素, 颜色相似性矩阵  $\mathbf{M}$  可以通过对颜色心理学的研究中获得 (Zhan H.J. et al 1995)。

## (4) 余弦距离 (Cosine distance)

余弦距离计算的是两个向量间方向的差异, 其定义如下 (Zhang D. S. 2002):

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{A}^T \mathbf{B} = |\mathbf{A}| \cdot |\mathbf{B}| \cos \theta \quad (2-14)$$

$$d_{\cos}(\mathbf{A}, \mathbf{B}) = 1 - \cos \theta = 1 - \frac{\mathbf{A}^T \mathbf{B}}{|\mathbf{A}| \cdot |\mathbf{B}|} \quad (2-15)$$

式中,  $|\mathbf{A}| = (\mathbf{A}^T \mathbf{A})^{\frac{1}{2}}$ ,  $|\mathbf{B}| = (\mathbf{B}^T \mathbf{B})^{\frac{1}{2}}$ 。

## (5) 相关系数

相关系数是一个可以用来表征两个向量之间线性关系紧密程度的量 (盛骤等 1989), 其数学描述如下:

$$\rho(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}} \quad (2-16)$$

采用相关系数, 两向量间的距离可表示如下:

$$d_{\rho} = 1 - \rho(\mathbf{A}, \mathbf{B}) \quad (2-17)$$

式中,  $\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i$ ,  $\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i$ 。在 (孙君顶等 2004a) 中, 采用相关系数来进行图像间的距离度量。



## (6) Kullback-Leibler 散度和 Jeffrey 散度

Kullback-Leibler (K-L) 散度定义为 (Kullback S. 1968)

$$d_{kl}(A, B) = \sum_{i=1}^n a_i \log \frac{a_i}{b_i} \quad (2-18)$$

式中,  $a_i \geq 0$ ,  $b_i \geq 0$ , 且  $\sum_{i=1}^n a_i = 1$ ,  $\sum_{i=1}^n b_i = 1$ 。K-L 散度的缺点是非对称并且对直方图柱值数敏感, 改进后的 Jeffrey 散度具有对称性和对噪声及直方图柱值数的健壮性, 其定义如下 (Puzicha J. 1997):

$$d_{jef}(A, B) = \sum_{i=1}^n \left[ a_i \log \frac{a_i}{m_i} + b_i \log \frac{b_i}{m_i} \right] \quad (2-19)$$

式中,  $m_i = \frac{a_i + b_i}{2}$ 。

(7)  $\chi^2$  距离

$\chi^2$  距离的定义如下 (Zhang D.S. et al 2003):

$$d_{\chi^2}(A, B) = \sum_{i=1}^n \frac{(a_i - m_i)^2}{m_i} \quad (2-20)$$

式中,  $m_i = \frac{a_i + b_i}{2}$ 。

## (8) EMD (Earth Mover's Distance) (Yossi R. et al 2000, 王文惠等 2000)

形象的解释 EMD 就是: 空间  $S$  中分布着  $m$  堆土  $p_i$ ,  $i=1, \dots, m$ , 每堆土的质量为  $w_{p_i}$ , 同时分布有  $n$  个土坑, 土坑的大小为  $q_j$ ,  $j=1, \dots, n$ , 每个土坑可以装土的质量为  $w_{q_j}$ , 即  $p = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$ ,  $q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$ 。把所有的土填到这些坑内, 做的功可表示为

$$\text{work}(p, q, f) = \sum_{i=1}^m \sum_{j=1}^n d(p_i, q_j) f_{ij} \quad (2-21)$$

式中,  $d(p_i, q_j)$  表示第  $i$  堆土到第  $j$  个坑的距离, 为了与图像的距离测度的区别, 称为基本距离。 $f_{ij}$  表示第  $i$  堆土运到第  $j$  个坑的土的质量。 $d(p_i, q_j) f_{ij}$  表示把第  $i$  堆土中质量为  $f_{ij}$  的土运到第  $j$  个坑所做的功。式 (2-21) 隐含的约束条件包括:

$$\textcircled{1} f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n$$

$$\textcircled{2} \sum_{j=1}^n f_{ij} \leq w_{p_i}, 1 \leq i \leq m$$

$$\textcircled{3} \sum_{i=1}^m f_{ij} \leq w_{q_j}, 1 \leq j \leq n$$

$$\textcircled{4} \sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left( \sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right)$$

条件①说明, 每次搬运土的质量大于零时才做工; 条件②说明, 从第  $i$  堆土运到各个坑土的质量一定不会大于该堆土的质量; 条件③说明第  $j$  个坑能够接受的土的质量, 一定不会大于该坑能够接受的土的最大量; 条件④说明从各堆土运到第  $j$  个坑土的质量, 一定不会大于该堆土的质量。从而, EMD 定义如下:

$$\text{EMD}(\mathbf{p}, \mathbf{q}) = \frac{\sum_{i=1}^m \sum_{j=1}^n d(p_i, q_j) f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (2-22)$$

从上述定义可以看出，我们不难把图像相似度度量的问题，转化为计算 EMD 的问题。当测量两幅图像的距离时，可以把一幅图像的特征矢量映射为土堆，把另一副图像的特征矢量映射为土坑，则两幅图像间的距离，就是把所有土填入土坑内，在选择最佳路径的条件下，做功的最小值。

#### (9) 编辑距离 (Edit Distance)

编辑距离又称为 Levenshtein 距离 (Levenshtein V. 1966)，被定义为从一个串变换到另一个串的最少插入、删除、替换操作。编辑距离广泛的应用于结构模式识别中 (Bunke H. et al 1999, Chen S.W. et al 1998)。

设  $A$  表示有限的字符集合， $\varepsilon$  表示一个空符号。编辑操作可以表示为  $a \rightarrow b, a \rightarrow \varepsilon, \varepsilon \rightarrow a$ ，其中  $a, b \in A$ ， $a \rightarrow b$  代表替换操作， $a \rightarrow \varepsilon$  代表删除操作， $\varepsilon \rightarrow a$  代表插入操作。设  $c(a \rightarrow b), c(a \rightarrow \varepsilon), c(\varepsilon \rightarrow a)$  分别表示各类操作的代价，设符号  $\mathbf{x} = x_1, x_2, \dots, x_n$ ， $\mathbf{y} = y_1, y_2, \dots, y_m$ ，则它们间的编辑距离的计算可表示为 4 个步骤 (Wagner R. et al 1974)：

步骤 1  $D(0, 0) = 0$

步骤 2 For  $j = 1, \dots, m$

$$D(0, j) = D(0, j-1) + c(\varepsilon \rightarrow y_j)$$

步骤 3 For  $i = 1, \dots, n$

$$D(i, 0) = D(i-1, 0) + c(x_i \rightarrow \varepsilon)$$

步骤 4 For  $i = 1, \dots, n$

For  $j = 1, \dots, m$

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + c(x_i \rightarrow y_j) \\ D(i-1, j) + c(x_i \rightarrow \varepsilon) \\ D(i, j-1) + c(\varepsilon \rightarrow y_j) \end{cases}$$

从上述的计算步骤可以看出，串  $x$  和  $y$  间的编辑距离  $d(\mathbf{x}, \mathbf{y}) = D(n, m)$ 。

#### (10) Hausdorff 距离

Hausdorff 距离是一种定义于两个点集上的最大-最小 (max-min) 距离，它主要用于测量两个点集的匹配程度，常用于二值图像或轮廓线的匹配技术中 (Huttenlocher D.P. et al 1993, 牛力丕等 2007)。给定两个有限点集  $\mathbf{P} = (p_1, p_2, \dots, p_m)$ ， $\mathbf{Q} = (q_1, q_2, \dots, q_n)$ ，则  $\mathbf{P}$  与  $\mathbf{Q}$  之间的 Hausdorff 距离定义为

$$D(\mathbf{P}, \mathbf{Q}) = \max\{d(\mathbf{P}, \mathbf{Q}), d(\mathbf{Q}, \mathbf{P})\} \quad (2-23)$$

式中  $d(\mathbf{P}, \mathbf{Q})$  为从点集  $\mathbf{P}$  到点集  $\mathbf{Q}$  的有向 Hausdorff 距离， $d(\mathbf{Q}, \mathbf{P})$  可作类推。

$$d(\mathbf{P}, \mathbf{Q}) = \max_{p \in \mathbf{P}} \min_{q \in \mathbf{Q}} \|p - q\| \quad (2-24)$$

$$d(\mathbf{Q}, \mathbf{P}) = \max_{q \in \mathbf{Q}} \min_{p \in \mathbf{P}} \|q - p\| \quad (2-25)$$

式中， $\|\cdot\|$  为定义在点集合上的某种距离范数。

这里，式 (2-23) 称为双向 Hausdorff 距离，是 Hausdorff 距离的最基本形式；式 (2-24) 及 (2-25) 分别称为从  $\mathbf{P}$  集到  $\mathbf{Q}$  集和从  $\mathbf{Q}$  集到  $\mathbf{P}$  集的单向 Hausdorff

距离。双向 Hausdorff 距离是单向距离  $d(P, Q)$  和  $d(Q, P)$  两者中的较大者，它度量了两个点集间的最大不匹配程度。

在上述的相似度度量方法中，没有任何一种方法可以适用于所有特征向量间的相似度度量。主要原因是，上述度量方法具有特征依赖的特点，不同的特征应该用不同的度量方法。例如，直方图相交法不适合于非直方图的特征；虽然二次式距离可以有效的度量颜色直方图的距离，但对其他特征向量的距离度量效果却没有采用欧氏距离度量的效果好 (Zhang D. S. 2002)。因此在进行图像检索时，难以证明上述相似度度量方法谁更具有优势，在具体的应用过程中，只有根据实际需要进行灵活选择。需要指出的是，在相似度度量的能力上，某些已知工作给出了一些结果，如文献 (Zhang D.S. 2002) 指出，除了颜色的直方图特征外，在进行特征向量间的相似度度量时，欧氏距离和城区距离相对于其他相似度度量方法具有更好的检索性能。

### 2.4.3 精确查询与近似查询

精确相似性查询是指，在给定特征向量和距离度量函数的前提下，获得精确的查询结果。近似相似性查询是指，在一定的误差允许下，获得的相似性查询结果。通过控制查询结果的不确切性，采用近似相似性查询可以提升查询性能。目前，近似近邻查询技术是高维数据检索的一个新的研究趋势，它提供了一种克服维数灾难现象的新的手段和方法。在高维情况下，绝大多数的索引结构对于大型多媒体数据库查询性能都很难达到预期效果。采用近似近邻查询，可以使用户在查询结果的精确性和查询时间上取得折中效果。但是，目前对于近似查询并没有严格的定义和统一的评价标准。研究者提出了许多近似查询的索引结构和搜索算法 (Ferhatosmanoglu H. et al 2001, Arya S. et al 1998, Ciaccia P. et al 2000, Pramanik S. et al 1999, Indyk P. et al 1998, 崔江涛 2005)，在这其中，具有严格定义的是  $\varepsilon$  近似近邻查询。

给定查询向量  $q$ ，一个近似系数  $\varepsilon (\varepsilon > 0)$ ，向量  $p$  是查询向量  $q$  的  $(1 + \varepsilon)$  近似近邻，如果满足下式：

$$d(p, q) \leq (1 + \varepsilon) d(p^*, q) \quad (2-26)$$

这时， $p^*$  是查询向量  $q$  的真正近邻。也就是说， $p$  与真正近邻的相对误差为  $\varepsilon$ 。更进一步，对于第  $k$  个  $(1 + \varepsilon)$  近似近邻与真正  $k$  近邻的相对误差也是  $\varepsilon$ ，即

$$d(p^k, q) \leq (1 + \varepsilon) d(p^{k*}, q) \quad (2-27)$$

$p^k$  和  $p^{k*}$  分别表示  $q$  的第  $k$  个近似近邻和真正近邻。

在多媒体信息检索应用中，特征向量和距离测度的选取都带有主观性或试探性，特征向量本身就只是多媒体内容的近似表示，而非精确表示，并且不能从数学上对对象之间的相似性进行严格的定义。由于多媒体信息所蕴含内容的丰富性和用户需求判断的主观性，很难找到一种精确的图像内容表示和度量方法。因此，传统在精确近邻搜索并不具有精确含义。在许多查询应用领域，都采用了人一机交互式的查询方式，搜索引擎需要根据用户对查询的反馈信息重新调整查询过程，比较典型的应用是基于相关反馈的多媒体检索系统。此类系统中，查询结果允许一定的不精确性，但是查询的响应时间是至关重要的。对用户来讲，可以容许一定的查询误差，来获得更好的查询性能。如何将近似查询方法和具体应用相结合，并利用模糊的定义方式达到有效查询的目的，是以后近似查询方法研究的重点。

## 2.5 性能评价准则

在进行图像检索时往往需要选择一种或多种有效的特征描述方法和相似度量方法，这就需要对不同的图像特征或特征组合以及不同的相似度量方法的检索效果进行全面的评价，比较不同方法的性能，找出最好的方法。但是，图像检索具有很强的主观性，同时也很难找到一个统一的图像测试库，因此，评价一个图像检索算法性能的优劣并不容易。这里列举了目前图像检索领域广泛应用的几个公认的图像检索算法评价准则（孙君顶，2005，Arnold W. M. et al, 2000, Manjunath B.S. et al, 2001, Mehtre B. M. et al, 1995, Hsieh I. S. et al, 2001, Jain A. K. et al, 1998, Lee H. Y. et al, 2003, 茹立云 等, 2003）。

### 2.5.1 精确度和检索率

精确度（precision）和检索率（recall）是目前在 CBIR 中应用最为广泛的一种评价准则。精确度的含义是，在一次查询过程中，系统返回的相关图像数目占所有返回图像数目的比例。如果在检索结果集合中，正确相关图像数目多，则精确度就高。检索率则指系统返回的查询结果中相关图像数目占图像库中所有相关图像数目（包括返回的和没有返回的）的比例。

设  $S$  为图像库中所有和查询图像相关的图像集， $R$  为所有检索到的图像集合， $s$  为一次查询中检索到的所有相关图像数目， $u$  为一次检索过程中检索到的不相关的图像数目， $v$  为图像库中和检索图像相关但在检索中未被检索到的图像的数目，这样精确度和检索率可表示为：

$$\text{recall} = P(R | S) = \frac{P(S \cap R)}{P(S)} = \frac{s}{s + v} \quad (2-28)$$

$$\text{precision} = P(S | R) = \frac{P(S \cap R)}{P(R)} = \frac{s}{s + u} \quad (2-29)$$

精确度和检索率越高，表明该检索系统的效果越好。通常，检索率和精确度是一对矛盾，当要求精度较高时，检索率较低，反之亦然。因此，一般的检索系统只要求在这两者之间达到一个最优的平衡点，就认为实现了较好的检索性能。

对于精确度和检索率来说，首先需要指导数据库中每一类图像的相似图像数目，因此对于大型图像库，尤其对于图像库中图像动态变化的图像库来说，要做到这一点比较困难。另外，要统计检索率和精确度，还需要用户在检索结果中标记出与示例图像相似的图像。由于用户的主观性，对于同一次查询来说，不同用户得到的精确度和检索率可能并不相同。

### 2.5.2 命中准确率

精确度和检索率需要用户在图像库中人工找出与查询图像相似的图像集，这将耗费大量的人工劳动，因此这种度量准则对于较小型的图像数据库比较合适。如果图像库测试集已经提前进行了分类，如 Corel Image Gallery 等类型的数据库，就可以简单的将每一个图像类别作为其中每一幅图像的相关图像，由此来度量算法的检索准确率。设图像  $q$  所在的

相关图像集为  $G$ ，图像检索算法自动输出了  $T$  幅相似图像，其中命中  $G$  的有  $n$  幅图像，此次检索的准确率定义为

$$P_T = \frac{n}{T} \quad (2-30)$$

由此，平均多个查询的检索准确率就可以度量算法的检索性能。

### 2.5.3 排序值评测法

设  $q$  是一幅查询图像， $g_1, g_2, \dots, g_n$  为图像检索算法检索到的与  $q$  相关的且从主观上认为相似的图像，设  $\text{rank}(g_i), i=1, 2, \dots, n$  为图像  $g_i$  在检索结果图像序列中对应的排序值，则下述两个指标可以有效的衡量算法的检索性能

$$r\text{-measure} = \frac{1}{n} \sum_{i=1}^n \text{rank}(g_i) \quad (2-31)$$

$$p\text{-measure} = \frac{1}{n} \sum_{i=1}^n \frac{i}{\text{rank}(g_i)} \quad (2-32)$$

其中，第一个指标定义了所有相关图像在检索结果中的平均排序，显然，此指标越小，检索算法的准确率越高。第二个指标定义了所有相关图像在靠前列出的紧密程度，该值越大表明检索结果越好，如果所有的相关图像都排在最前面，则此指标取值为1。

### 2.5.4 ANMRR

ANMRR (Average Normalized Modified Retrieval Rank) 是 MPEG-7 推荐的一种性能评价方法。设  $N(q_i) (i=1, 2, \dots, Q)$  表示图像库中与图像  $q_i$  相似的所有图像数目， $M = \max\{N(q_1), N(q_2), \dots, N(q_Q)\}$ ， $K = \min\{4N(q_i), 2M\}$ ，设与例子图像相似的图像在检索结果序列中所处的位置为

$$\text{rank}(k) = \begin{cases} k; & k \leq K \\ K+1; & k > K \end{cases} \quad (2-33)$$

从而，ANMRR 定义如下

$$\text{ANMRR} = \frac{1}{Q} \sum_{i=1}^Q \frac{\sum_{k=1}^{N(q_i)} \frac{\text{rank}(k)}{N(q_i)} - 0.5 - 0.5 * N(q_i)}{K + 0.5 - 0.5 * N(q_i)} \quad (2-34)$$

由式 (2-34) 可知，ANMRR 的取值越小则该算法的检索性能越好。

应当指出，除了主观因素，图像检索的准确率与图像数据库有很大的关系，即使用同一算法，在不同的数据库中根据相同的评价方法计算出的检索准确率，都可能会存在较大的差异。

### 2.5.5 前 $N$ 个结果的正确率与检索率

前  $N$  个结果的正确率描述如下：设  $R$  为某一具有特定语义含义的图像集合， $q_i \in R$  为任一检索示例图像，在一次检索过程中，若系统返回  $N$  个结果（记为  $I_1, I_2, \dots, I_j, \dots, I_N$ ），则正确率  $P_N(q_i)$  定义为

$$P_N(q_i) = \sum_{j=1}^N \frac{\phi(I_j, R)}{N} \quad (2-35)$$

式中,  $\phi(I_j, R) = \begin{cases} 0, & \text{if } I_j \notin R \\ 1, & \text{if } I_j \in R \end{cases}$ , 那么, 对于所有测试样例检索图像集得到的平均正确率可表示为

$$P_N = \sum_{i=1}^K \frac{P_N(q_i)}{K} \quad (2-36)$$

式中,  $K$  表示测试样例检索图像集中的图像数目。该正确率的定义简单的说, 就是在返回的前  $N$  个结果中的正确的比例。

前  $N$  个结果的检索率定义为

$$R_N(q_i) = \sum_{j=1}^N \frac{\phi(I_j, R)}{\|R\|} \quad (2-37)$$

$\|R\|$  表示图像集  $R$  中的图像数目, 那么, 对于所有测试样例检索图像集得到的平均检索率可表示为

$$R_N = \sum_{i=1}^K \frac{R_N(q_i)}{K} \quad (2-38)$$

## 参 考 文 献

- [1] Adoram M., Lew M.S., 1999. IRUS: image retrieval using shape. IEEE International Conference on Multimedia Computing Systems. 2: 597-602.
- [2] Arnold W.M., Marcel W., Simone S., et al, 2000. Content-based image retrieval at the end of the early years. IEEE Trans. On PAMI, 22(12): 1349-1379.
- [3] Arya S., Mount D.M., Netanyahu N.S., et al., 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. Journal of ACM, 48(6): 891-923.
- [4] Bober, M., MPEG-7 visual shape descriptors, IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(6): 716-719.
- [5] Bunke H., Zumbühl M., 1999. Acquisition of 2D shape models from scenes with overlapping objects using string matching. Pattern Analysis and Applications, 2: 2-9.
- [6] Chang, S.F., Puri, A., Sikora, T. et al, 2001. Overview of the MPEG-7 standard. IEEE Trans. Circuit System Video Technology, 11: 688-695.
- [7] Chen S.W., Tung S.T., Fang C.Y., et al., 1998. Extended attributed string matching for shape recognition. Computer Vision and Image Understanding, 70: 36-50.
- [8] Ciaccia P., Patella M., 2000. PAC nearest neighbor queries: Approximate and controlled search in high-dimensional and metric spaces. In Proceedings of the 16th International Conference on Data Engineering, San Diego, California USA, IEEE Computer Society.
- [9] Cinque L., Levialdi S., Olsen K., et al, 1999. Color-based image retrieval using spatial-chromatic histograms. In Multimedia Computing and System, IEEE Intl Conf. Vol 2, 969-973.
- [10] 崔江涛. 高维索引技术中向量近似方法研究. 博士学位论文, 西安电子科技大学, 2005.
- [11] Dimai A., 1997. Spatial encoding using differences of global features. In Proc. of SPIE storage and retrieval for image and video database IV, 3022, 352-360.
- [12] Eakins J.P., Graham M.E., 1999. Content-based image retrieval (A report to the JISC Technology

- Applications Programme). Newcastle upon Tyne, United Kingdom: University of Northumbria at Newcastle, Institute for Image Data Research.
- [13] Ebrahimi T., Abdeljaoued Y., Figureas R.M., et al, 2001. MPEG-7 camera, in: Proceedings of the International Conference on Image Processing, Thessaloniki, 3: 600-603.
- [14] Faloutsos C., Barber R., Flickner M., et al, Efficient and effective querying by image content. *Intelligent information systems*, 3(1): 231-262, 1994.
- [15] Farzin Mokhtarian, Alan K. Mackworth, 1992. A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8): 789-805.
- [16] Ferhatosmanoglu H., Tuncel E., Agrawal D., et al, 2001. Approximate nearest neighbor searching in multimedia databases. In *Proc. 17th IEEE Int. Conf. Data Eng. (ICDE)*, 503-511.
- [17] Gimelfarb G., 1999. Modeling image textures by Gibbs random fields. *Pattern Recognition Letters*, 20(11): 1123-1132.
- [18] Granlund G.H., 1972. Fourier processing for hand print characters recognition. *IEEE Trans. Computers*, c-21: 195-201.
- [19] Gudivada V.N., Raghavan V.V., 1995. Content-based image retrieval system. *IEEE Computer*, 28(9): 18-22.
- [20] Hadid A., Pietikäinen M. and Li SZ, 2006. Boosting spatio-temporal LBP patterns for face recognition from video. *Proc. Asia-Pacific Workshop on Visual Information Processing (VIP2006)*, Beijing, China, 75-80.
- [21] Hafner J., Sawhney H., Equitz W., et al, 1995. Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 17(7): 729-736.
- [22] Hanjalic A. 2001. Video and image retrieval beyond the cognitive level: the needs and possibilities, *Proceedings of SPIE Storage and Retrieval for Media Databases*, San Jose, CA, 130-140.
- [23] Haralick R. M., Shanmugam K., 1973. Texture features for image classification. *IEEE Trans. On system, man and cybernetics*, SMC-3(6): 610-621.
- [24] He D.C., Wang L., 1991. Texture features based on texture spectrum. *Pattern Recognition*, 24 (5): 391-399.
- [25] 何清法, 李国杰. 综合分块主色和相关反馈技术的图像检索方法. *计算机辅助设计与图形学学报*, 2001, 13(10): 912-917.
- [26] Hou Z., Qin Y.W., 2003. The study of fractal correlation method in the displacement measurement and its application. *Optics and Lasers in Engineering*, 39(4): 465-472.
- [27] Hsieh I.S., Fan K C., 2001. Multiple classifiers for color flag and trademark image retrieval. *IEEE Trans. On Image Processing*, 10(6): 938-950.
- [28] Hsu W., Chua T. S., Pung H. K., 1995. An integrated color-spatial approach to content-based image retrieval. In: *Proc ACM Multimedia'95 Conference*. San Francisco, 305-313, 1995.
- [29] Hu M.K., 1962. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 8(2): 179-187.
- [30] Huang J., 1998. Color-Spatial Image Indexing and Applications. Phd thesis, Cornell University, New York, USA.
- [31] 黄元元, 郭丽, 杨静宇. 基于目标区域颜色与纹理特征的图像检索. *南京理工大学学报*, 2003, 27(3): 286-289.
- [32] Huang T.S., Mehrotra S., Ramlchandran K., 1996. Multimedia analysis and retrieval system (MARS) project, In *Proc. of 33rd Annual Clinic on Library Application of Data Processing Digital Image Access and Retrieval*.
- [33] 黄祥林, 沈兰荪. 基于内容的图像检索技术研究. *电子学报*, 2002, 30(7): 1065-1071.
- [34] Huttenlocher D.P., Klanderman G.A., Rucklidge W.J., 1993. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(9): 850-863.

- [35] Iivarinen J., Visa A., 1996. Shape recognition of irregular objects, In: Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, SPIE, 25-32.
- [36] Indyk P., Motwani R., 1998. Approximate nearest neighbors: Towards removing the curse of dimensionality. Proc. 30th ACM Symposium on Theory of Computing, New York, ACM, 604-613.
- [37] Jain A.K., Vailaya A., 1998. Shape-based retrieval: a case study with trademark image databases. Pattern Recognition, 31(9): 1369-1390.
- [38] Jia Li, Gray R.M., 2000. Context-Based Multiscale Classification of Document Images using Wavelet Coefficient Distributions. Image Processing, IEEE Transactions, 9(9): 1604-1616.
- [39] John ZM, 2000. An Information theoretic approach to content based image retrieval. Louisiana State University and Agricultural and Mechanical College, Phd. Thesis.
- [40] Kamila N.K., Mahapatra S., Nanda S., 2005. Invariance image analysis using modified Zernike moments. Pattern Recognition Letters 26: 747-753.
- [41] Khotanzad, A., Hernandez, O.J., 2003. Color image retrieval using multispectral random field texture model and color content features, Pattern Recognition, 36(8): 1679-1694.
- [42] Koenen, R. Pereira, F., 2000. MPEG-7: A standardised description of audiovisual content. Signal Process. Image Commun. 16(1): 5-13.
- [43] Kuan J.P., Joyce D.W., 1999. Lewis P. H., Texture content based retrieval using text descriptions. SPIE, 3656: 75-85.
- [44] Kullback S., 1968. Information theory and statistics, Dover, New York, NY.
- [45] Lee H.Y., Lee H.K., Ha Y.H., 2003. Spatial color descriptor for image retrieval and video segmentation. IEEE Trans. On Multimedia, 5(3): 358-367.
- [46] Levenshtein V., 1966. Binary codes capable of correcting deletions, Insertions and Reversals. Soviet Physics Doklady, 10(8): 707-710.
- [47] Li C.T., 2003. Multiresolution image segmentation integrating Gibbs Sampler and region merging algorithm. Signal Processing, 83(1): 67-78.
- [48] Li J., Wang J.Z., Wiederhold G., 2000. IRM: integrated region matching for image retrieval, Proceeding of ACM Multimedia.
- [49] 李向阳, 庄越挺, 潘云鹤. 基于内容的图像检索技术与系统. 计算机研究与发展, 2001, 38(3): 344-354.
- [50] Lim S., Lu G. J., 2003. Spatial statistics for content -based image retrieval. Proceedings of the International Conference on Information Technology: Computers and Communications(ITCC).
- [51] Liu F., Picard R., 1996. Periodicity, directionality and randomness: wold features for image modeling and retrieval. IEEE-PAMI, 18(7): 722-733.
- [52] 刘忠伟. 利用局部累加直方图进行彩色图像检索. 中国图象图形学报, 1998, 3(7): 533-537.
- [53] Lowitz G.E., 1983. "Can a local histogram really map texture information?", Pattern Recognition. 16(2): 141-147.
- [54] Lu G.J., Sajjanhar A., 1999. Region-based Shape Representation and Similarity Measure Suitable for Content-based Image Retrieval. Multimedia System, 7(2): 165-174.
- [55] Lu G.J., Sajjanhar A., 1998. On Performance Measurement of Multimedia Information Retrieval Systems . In Proc . of International Conference on Computational Intelligence and Multimedia Applications, Australia, 781-787.
- [56] 卢汉清. 图像视频信号的浏览与检索. 中国图象图形学报, 2000, 5(B)(1): 5-9.
- [57] Manjunath B.S., Phillippe S., Thomas S., 2002. Introduction to MPEG-7: multimedia content description interface, John Wiley & Sons, Inc., New York, NY.
- [58] Manjunath B.S., Ma W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE



- Transactions on Pattern Analysis and Machine Intelligence, 18(8): 837-842.
- [59] Marko H., Matti P., Cordelia S., 2009. Description of interest regions with local binary patterns. *Pattern Recognition* 42, 425-436.
  - [60] Mehtre B.M., Kankalllalli M.S., Narasimhalu A.D., 1995. Color matching for image retrieval. *Pattern Recognition Letters*, 16: 325-331.
  - [61] Michael D., 2001. Wavelet features for color image classification. *Imaging and Geospatial Information Society*. Orlando.
  - [62] Minghong P., Mandal M.K., Basu A., 2005. Image retrieval based on histogram of fractal parameters. *IEEE Transactions on Multimedia*, 7(4): 597-605.
  - [63] Mokhtarian F., Mackworth A.K., 1992. A theory of multiscale: curvature based shape representation for planar curves. *IEEE Trans. Patern Analysis Machine Intelligence*, 14 (8): 789-805.
  - [64] Mokhtarian F., Abbasi S., Kittler J., 1996a. Robust and efficient shape indexing through curvature scale space, *Proceedings of the British Machine Vision Conference*, Edinburgh, UK, 53-62.
  - [65] Mokhtarian F., Abbasi S., Kittler J., 1996b. Efficient and robust retrieval by shape content through curvature scale space, *International Workshop on Image Databases and Multimedia Search*, Amsterdam, The Netherlands, 35-42.
  - [66] 木拉提·哈米提, 刘伟, 童勤业. 纹理谱直方图与潜在语义标引在图像检索中的应用. *科技通报*, 2006, 22(3): 389-394.
  - [67] Niblack W., Barber R., Equitz W., et al, 1993. The QBIC project: querying images by content using color, texture, and shape. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, 2-3 February '93, San Jose, CA, pages 173-187.
  - [68] 牛力丕, 毛士艺, 陈炜. 基于 Hausdorff 距离的图像配准研究. *电子与信息学报*, 2007, 29(1): 35-38.
  - [69] Ojala T., Pietikäinen M. and Harwood D., 1996. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29(1): 51-59.
  - [70] Ojala T. Pietikäinen M., 1998. Nonparametric multichannel texture description with simple spatial operators. *Proc. 14th International Conference on Pattern Recognition*, Brisbane, Australia, 1052-1056.
  - [71] Ojala T., Pietikäinen M., 1999. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3): 477-486.
  - [72] Ojala T., Pietikäinen M., Mäenpää T., 2000. Gray scale and rotation invariant texture classification with local binary patterns. In: *Computer Vision, ECCV 2000 Proceedings*, Lecture Notes in Computer Science 1842, Springer, 404-420.
  - [73] Ojala T., Valkealahti K., Oja E., Pietikäinen M., 2000. Texture discrimination with multidimensional distributions of signed gray level differences. *Proc. 4th Asian Conference on Computer Vision*, Taipei, Taiwan, 2: 1082-1088.
  - [74] Ojala T., Pietikainen M. and Maenpaa T., 2002. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24: 971-987.
  - [75] Paget R., 2004. Srtong Markov random field model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3): 408-413.
  - [76] Pass G, Zabin R, Miller J., 1996. Comparing Images Using Color Coherence Vectors. In *ACM International Conference on Multimedia*. Boston, MA, 65-73, 1996.
  - [77] Pentland A, Picard R, Sclaroff S., 1994. Photobook: tools for content-based manipulation of image database. *Proc. SPIE*, vol. 2185, 34-47.
  - [78] Persoon E. Fu K.S., 1997. Shapes description using fourier descriptors. *IEEE Trans. Systems. Man and Cybernetics*, SMC-7: 170-179.

- [79] Pramanik S., Alexander S., Li J., 1999. An efficient searching algorithm for approximate nearest neighbor queries in high-dimensions. In ICMCS 1999, IEEE International Conference on Multimedia Computing and Systems, Florence, Italy. IEEE Computer Society.
- [80] Puzicha J., Hufmann T., Buhmann J., 1997. Non-parametric similarity measure for unsupervised texture segmentation and image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 267-272.
- [81] Qiu G.P., Lam K.M., 2003. Frequency layered color indexing for content-based image retrieval. IEEE Transactions on Image Processing. 12(1): 102-113.
- [82] Randen T., Husoy J.H., 1999. Filtering for texture classification: A comparative study. IEEE Trans. Pattern Analysis and Machine Intelligence. 21(4): 291-310.
- [83] Rao A., Srihari R., Zhang Z., 1999. Spatial Color Histogram for Content-Based Retrieval, In 11th IEEE International Conference on Tools with AI, 183-186.
- [84] Rellier G., Descombes X., Falzon F., 2004. Texture feature analysis using a Gauss-Markov model in hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 42(7): 1543-1551.
- [85] 茹立云, 彭潇, 苏中等. 基于内容图像检索中的特征性能评价. 计算机研究与发展, 2003, 40(11): 1566-1570.
- [86] Santini S., Jain R., 1996. Similarity match. IEEE Trans. On PAMI, 18(9): 946-958.
- [87] Sciascio E.D., Mingolla G., Mongiello M., 1999. Content-based image retrieval over the web using query by sketch and relevance feedback. In: Proc of VISUAL'99, 123-130.
- [88] 盛骤, 谢式千, 潘承毅. 概率论与数理统计. 北京: 高等教育出版社, 1989.
- [89] Sebe N., Lew M.S., Huijsmans D.P., 1999. Multi-scale sub-image search. In Proc. of 7th ACM Int. Conf. On Multimedia (Part 2), 79-82.
- [90] Sebe N., Lew M. S., 2001. Texture features for content-based retrieval. In: Principles of Visual Information Retrieval. Lew M S, ed. Springer. 3, 51-85.
- [91] 施智平, 胡宏, 李清勇等. 一种快速有效的图像纹理谱描述子, 计算机辅助设计与图形学学报, 2004, 16(12): 1703-1707.
- [92] 帅勤. 基于MPEG-7的纹理和颜色特征的图像检索技术. 硕士学位论文, 广西大学, 2006.
- [93] Stehling R.O., Nascimento M.A., Falcao A.X., 2000. On 'shapes' of colors for content-based image retrieval. In the ACM Multimedia Conference. Los Angeles, USA, 171-174.
- [94] Stricker M., Orengo M., 1995. Similarity of color images. In: Proceedings of SPIE Storage and Retrieval for Image and Video Database, 2420: 381-392.
- [95] Stricker M., Dimai A., 1997. Spectral covariance and fuzzy regions for image indexing. Machine Vision and Applications, Vol 10, 66-73.
- [96] 孙劲光. 基于颜色特征的图像数据管理模式研究. 博士学位论文, 辽宁工程技术大学, 2006.
- [97] 孙君顶, 崔江涛, 毋小省等. 基于颜色和形状特征的彩色图像检索方法. 中国图象图形学报, 2004a, 9(7): 820-827.
- [98] 孙君顶, 毋小省, 周利华. 基于信息熵的图像检索. 西安电子科技大学学报, 2004b, 31(2): 223-228.
- [99] 孙君顶. 基于内容的图像检索技术研究. 博士学位论文, 西安电子科技大学, 2005.
- [100] 孙君顶, 周利华. 一种改进的基于熵的图像检索算法. 红外技术, 2005, 27(1): 45-48.
- [101] Sun J.D., 2006. Image Retrieval Based on Improved Entropy and Moments, International Conference on Intelligent Information Hiding and Multimedia Signal Processing, USA.
- [102] Sun J.D., Wu X.S., 2006a. Chain code distribution-based image retrieval. In: Intelligent Information Hiding and Multimedia Signal Processing, 139-142.
- [103] Sun J.D., Zhang X.M., Cui J.T., et al, 2006b. Image retrieval based on color distribution entropy, Pattern

- Recognition Letters, 27(10): 1122-1126.
- [104] Sun J.D., Wu X.S., 2007. Shape retrieval based on the relativity of chain codes. International Workshop Proceedings, Multimedia Content Analysis and Mining (MCAM'2007), Weihai, China, 76-84.
- [105] Sun J.D., Zhang Z.S., 2008. Shape retrieval based on combination moment invariants, Proceedings of Information Technology and Environmental System Science, Volume 3, 301-305.
- [106] 孙君顶. 基于链码分布特征及相关性的轮廓描述与检索. 光电子.激光, 2008, 19(8): 1112-1115.
- [107] Sung M.K., Soo J.P., Chee S.W., 2007. Image retrieval via query-by-layout using MPEG-7 visual descriptors. ETRI Journal, 29(2): 246-248.
- [108] Swain M. J., Ballard D. H., 1991. Color indexing. Intl. J. on Computer Vision, 7(1): 11-32.
- [109] Teague M.R., 1980. Image analysis via the general theory of moments, J. Opt. Soc. Am. 70 (8): 920-930.
- [110] Tumara H., Mori S., Yamawaki T., 1978. Texture features corresponding to visual perception. IEEE-SMC, 8(6): 460-473.
- [111] Yang L., Algreysen F., 1994. Fast Computation of Invariant Geimetrie Moments: ANew Method Giving Correct Results, Proc. IEEE Int. Conf. On Image Processing.
- [112] 杨翔英, 章毓晋. 小波轮廓描述符及在图像查询中的应用. 计算机学报, 1999, 22(7): 752-757.
- [113] Yoo H.W., Jang D.S., NA Y.K., 2002. An efficient indexing structure and image representation for content-based image retrieval. IEICE Trans INF&SYST, 1390-1398.
- [114] Yossi R., Carlo T., Leonidas G., 2000. The Earth Mover's Distance as a Metric for Image Retrieval. International Journal of Computer Vision, 40(2): 99-121, 2000.
- [115] Young D.C., Sang Y.S., Nam C.K., 2003. Image Retrieval using BDIP and BVLC Moments. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13 (9): 951-957.
- [116] Wagner R. and Fischer M., 1974. The string-to-string correction problem. Journal of the ACM, 21: 168-173.
- [117] 万华林, Chowdhury Morshed U, 胡宏等. 图像纹理特征及其在 CBIR 中的应用. 计算机辅助设计与图形学学报, 2003, 15 (2): 195-199.
- [118] Zhan H.J., Zhong D., 1995. A scheme for visual feature based image retrieval. In Proc SPIE Storage and Retrieval for Image and Video Database, 2420: 36-46.
- [119] Zhang D.S., 2002. Image Retrieval Based on Shape, PhD Thesis, Monash University.
- [120] 王惠锋, 孙正兴, 王箭. 语义图像检索研究进展. 计算机研究与发展, 2002, 39(5): 513-523.
- [121] 王文惠, 王展, 周良柱等. 一种测量图像相似性的新方法. 国防科技大学学报, 2000, 22(6): 87-90.
- [122] 王郑耀, 2006. 图像的视觉信息表象理论及其数学基础. 硕士学位论文, 西安交通大学.
- [123] 汪祖媛, 梁栋, 李斌. 基于树状小波分解的纹理图像检索. 中国图象图形学报, 2001, 6A(11): 1065-1069.
- [124] 吴冬升, 吴乐南, 黄波. 基于小波模糊聚类的均质纹理和非均质纹理图像检索. 中国图象图形学报, 2003, 8(12): 1400-1405.
- [125] Zahnand C.T., Roskies R.Z., 1972. Fourier descriptors for plane closed curves. IEEE Trans. Computers, c-21: 269 -281.
- [126] Zhang D.S., 2002. Image retrieval based on shape, PhD Thesis, Monash University, March.
- [127] Zhang D.S., Lu G.J., 2003. Evaluation of similarity measurement for image retrieval, IEEE Int. Conf. Neural Networks & Signal Processing, NanJing, China, 928-931.
- [128] 章毓晋. 图像分割. 北京: 科学出版社, 2001.
- [129] 章毓晋. 基于内容的视觉信息检索. 北京: 科学出版社, 2003.
- [130] 赵珊, 孙君顶, 周利华. 基于方块编码的图像纹理特征提取及检索算法. 光电子激光, 2006, 17(8): 1014-1017.
- [131] 赵珊. 基于内容的图像检索关键技术研究. 博士学位论文, 西安电子科技大学, 2007.

- [132] 赵珊, 周利华. DCT 压缩域中基于纹理和形状的图像检索算法. 西安电子科技大学学报, 2007, 34(3): 402-408.
- [133] Zhou X.S., Thomas S.H., 2001. Edge-Based Structural Features for Content-Based Image Retrieval. Pattern Recognition Letters, 22(5): 457-468.
- [134] 周向东. 图像数据库中的关键技术研究. 博士学位论文, 复旦大学, 2003.
- [135] 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索. 北京: 清华大学出版社, 2002.

## 第3章 基于颜色特征的图像检索

颜色特征是在图像检索中应用最为广泛的视觉特征。一方面，由于颜色往往和图像中所包含的物体或场景十分相关；另一方面，相对于纹理和形状等低层特征，颜色特征对图像本身的尺寸、方向、视角、平移、旋转等的依赖性较小，具有较强的鲁棒性。因此，基于颜色的图像表示方法自然就成为一种主要的图像索引技术，并得到相当广泛和深入的研究。本章将从颜色空间与颜色量化入手，介绍常用的颜色特征提取及描述方法。

### 3.1 引言

图像特征的提取与表达是基于内容的图像检索技术的基础。从广义上讲，图像的特征包括基于文本的特征（如关键字、注释等）和基于视觉的特征（如色彩、纹理、形状、对象表面等）两类。视觉特征又可分为通用的视觉特征和领域相关的视觉特征。前者用于描述所有图像共有的特征，与图像的具体类型或内容无关，主要包括色彩、纹理和形状；后者则建立在对所描述图像内容的某些先验知识（或假设）的基础上，与具体的应用紧密相关。

颜色特征是在图像检索中应用最为广泛的视觉特征，也是人识别图像的主要感知特征。主要原因在于，颜色往往和图像中所包含的物体或场景十分相关。自然界中，同一类物体通常有相同或相近的颜色特征，不同类的物体则可能表现为不同的颜色特征，所以，颜色经常可以作为区分不同物种最为简单有效的一种手段。与其他的视觉特征相比，颜色特征对图像本身的尺寸、方向、视角的依赖性较小，具有较高的鲁棒性。因此，大多数图像检索系统都将颜色特征作为图像检索的主要检索手段。

基于颜色特征的图像检索系统中涉及若干问题。首先，需要选择合适的颜色空间来描述颜色特征；其次，采用一定的量化方法将颜色特征表达为向量的形式；再次，以一定的方式来描述颜色特征；最后，还要定义一种相似度（距离）标准来衡量特征间的相似性。因此，本章首先讨论了基于内容图像检索中常用的颜色空间及颜色量化的手段，在此基础上，分别介绍了目前常用的颜色特征描述方法。

### 3.2 颜色空间

从数字图像中提取颜色特征，依赖于对数字图像中颜色的表示和颜色理论的理解。颜色空间对于相关颜色以数字形式表示是一个很重要的成分，在不同颜色空间之间的转换和颜色信息的量化是给定特征提取方法的决定因素。

#### 3.2.1 颜色基础

通常，被人类感觉到的颜色称为光，它是一段连续的电磁波，波长  $400\sim 780\text{nm}$ 。神经生理学实验发现，在视网膜上存在三种不同的颜色感受器，它们是三种不同的锥体细胞，

每一种有自己独特的光谱灵敏度及其灵敏范围，感受到的颜色也就是我们通常所说的三基色，即红色、绿色和蓝色，它们相应的波长分别为 700nm、546.1nm、435.8nm (Kenneth.R.Castleman 1998)。利用三基色叠加可产生光的三补色——品红（即红加蓝），蓝绿（即绿加蓝），黄（即红加绿）。按一定的比例混合三基色，或一个补色与相对的基色混合就可以产生白色。

另外，颜色和彩色严格来说并不等同。颜色可分为无彩色和有彩色两大类。无彩色指白色、黑色和各种深浅程度不同的灰色。以白色为一端，通过一系列从浅到深排列的各种灰色，到达另一端的黑色，这些可以组成一个黑白系列。彩色则指除去上述黑白系列以外的各种颜色。不过我们通常所说的颜色一般指彩色。

自然界中常见的各种色光都可以用这三原色按照不同比例混合得到。这就是三原色原理。该原理是 T. Young 在 1802 年提出的，三原色原理指出 (姚敏等 2006)：

(1) 自然界中的可见颜色都可以用三种原色按一定比例混合得到；反之，任意一种颜色都可以分解为三种原色。

(2) 作为原色的三种颜色应该互相独立，即其中任何一种都不能用其他两种混合得到。

(3) 三原色之间的比例直接决定混合色调的饱和度。

(4) 混合色的亮度等于各原色亮度之和。

从光学物理学的角度出发，颜色的三个特性分别为主波长 (Dominant Wavelength)、纯度 (Purity) 和明度 (Luminance)。主波长是产生颜色光的波长，对应于视觉感知的色调；光的纯度对应于饱和度；而明度就是光的亮度 (Kenneth.R.Castleman 1998, 章毓晋 1999, 汤海纓等 1999)。

从心理学和视觉的角度出发，颜色具有三个特性：色调 (Hue)、饱和度 (Saturation) 和亮度 (Value)。所谓色调，是一种颜色区别于其他颜色的决定因素，我们平常所说的红、绿、蓝、紫等颜色，实际上指的是色调；饱和度是指颜色的纯度，鲜红色饱和度高，而粉红色的饱和度低；明度是与所观察物体明亮程度相关的视觉特征，由物体漫反射或漫透反射光的程度决定。关于色调、饱和度和亮度的概念，我们将在讨论 HSV 颜色空间时详细论述。

对彩色图像进行研究，必须在特定的颜色空间中进行。实际应用中常用的颜色空间很多，在图像检索中，主要采用的颜色空间有 RGB、HSV、YCrCb、CIEL\*a\*b\* 和 CIEL\*u\*v\* 等 (孙君顶 2005, 章毓晋 1999, 汤海纓等 1999, Mathias E. et al 1998)，下面我们将分别介绍。

### 3.2.2 RGB颜色空间

RGB 颜色空间是图像处理中最基础、最常用的颜色空间。现有的图像采集设备最初采集到的颜色信息是 RGB 值，颜色显示设备最终使用的也是 RGB 值，图像处理中使用的其他颜色空间也是通过 RGB 颜色空间转换来的。

RGB 颜色空间的三维空间包括 R、G、B 三个坐标轴，如图 3.1 所示。我们感兴趣的部分是个立方体，原点对应黑色，离原点最远的顶点对应白色，立方体与三个坐标轴的交点对应于三基色——红色、绿色和蓝色，剩余的

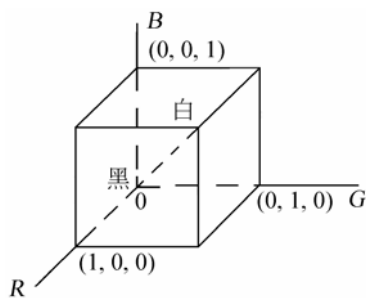


图 3.1 RGB 颜色模型

3 个顶点对应于三补色——品红（即红加蓝）、蓝绿（即绿加蓝）和黄（即红加绿）。在这个模型中，从黑到白的灰度值分布在从原点到离原点最远顶点间的连线上，而立方体内其余各点对应的不同颜色可以用该点到原点的向量来表示。方便起见，可将立方体归一化为单位立方体，因此  $R$ 、 $G$ 、 $B$  的值都在区间  $[0, 1]$  中（章毓晋 1999）。根据这个模型，每幅彩色图像包括 3 个独立的基色平面，或者说可分解到 3 个平面上。

色觉的产生需要光源的光通过反射或透射的方式传递到眼睛，刺激视网膜细胞，引起神经信号传输到大脑，之后，人脑对此加以解释产生视觉。设为组成某颜色  $C$  所需的 3 个刺激量分别用  $X$ 、 $Y$  和  $Z$  表示，与  $R$ 、 $G$  和  $B$  有如下关系（章毓晋 1999）：

$$\begin{aligned} X &= 0.49R + 0.310G + 0.200B \\ Y &= 0.177R + 0.812G + 0.011B \\ Z &= 0.000R + 0.010G + 0.990B \end{aligned} \quad (3-1)$$

对白光，有  $X=1$ ， $Y=1$ ， $Z=1$ 。设每种刺激量的比例系数为  $x, y, z$ ，则有  $C = xX + yY + zZ$ 。比例系数  $x, y, z$  也称为色系数，其定义为：

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z} \end{aligned} \quad (3-2)$$

可以看出  $x + y + z = 1$ 。

RGB 颜色空间的主要缺点是不直观，从 RGB 值中很难知道该值所表示颜色的认知属性，因此，RGB 颜色空间不符合人对颜色的感知心理（魏宝刚等 1999）。另外，RGB 颜色空间是最不均匀的颜色空间之一，两种颜色之间的知觉差异不能采用该颜色空间中两个颜色点之间的距离来表示（徐旭等 1999）。

### 3.2.3 HSV颜色空间

HSV 颜色空间是一种面向视觉感知的颜色模型。上面提到的 RGB 颜色模型是从硬件的角度提出的，并不能很好的与人眼的视觉特征相匹配。从心理学和视觉的角度出发，人眼的色彩知觉主要包括 3 个要素：色调、饱和度和亮度（章毓晋 1999）。

色调是指光的颜色，它与混合光谱中主要光波长相联系的，如红、橙、黄、绿、青、蓝、紫等颜色分别表示不同的色调。就波长的意义上讲，不同波长的光呈现不同的颜色，就是指色调不同。

饱和度是指彩色的深浅程度，即与一定色调的纯度相关。饱和度高表示颜色深，如深红，饱和度低，则颜色浅，如浅红。饱和度的深浅与色光中白光成分的多少有关，一种纯彩色光中加入的白光成分越少，其饱和度就越高；反之，白光成分越多，饱和度就越低。饱和度反映了某种色光被冲淡的程度。

亮度是指人眼感受到的光的明暗程度，与物体的反射率成正比。如果是无彩色就是指只有亮度一个维的变化。对彩色来说，颜色中掺入白色越多就越明亮；掺入黑色越多亮度就越小。

用一个三维空间纺锤体可以将色调、饱和度和亮度表示出来，如图 3.2 所示。其中立

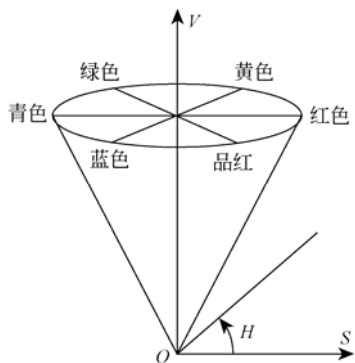


图 3.2 HSV 颜色模型

体的竖直轴代表黑白系列亮度的变化，圆周上各点代表不同的色调，从圆周向圆心过渡表示饱和度逐渐降低。HSV 颜色空间便是这样一个可以用于描述颜色的色调、饱和度和亮度的彩色模型，该模型反映了人观察彩色的方式 (Kenneth.R.Castleman 1998, 魏宝刚等 1999)。

HSV 颜色空间具有两大特点：其一，亮度分量与图像的彩色信息无关；其二，色调和饱和度分量与人感受颜色的方式是紧密相连的。这些特点使 HSV 颜色空间非常适合于借助人的视觉系统来感知颜色特性的图像处理算法 (Kenneth.R.Castleman 1998)。HSV 颜色空间直接对应于人眼色彩视觉特征的三要素，通道之间各自独立，可以被独立感知各颜色分量的变化，其中色调尤其影响着人的视觉判断。因此，在基于内容的图像检索中，应用这种彩色模型会更适合用户的视觉判断。

由于数字图像一般均采用 RGB 颜色模型来显示，因此在进行处理时，需要将颜色值由 RGB 空间转换至 HSV 空间。给定 RGB 颜色空间中的值  $(R, G, B)$ ， $R, G, B \in [0, 1]$ ，则转换到 HSV 空间的  $H$ 、 $S$ 、 $V$  值计算如下 (孙兴华 2001, 章毓晋 1999)：

$$V = \frac{1}{\sqrt{3}} [R + G + B] \quad (3-3)$$

$$S = 1 - \frac{\sqrt{3}}{V} \min(R, G, B) \quad (3-4)$$

$$H = \begin{cases} \theta & G \geq B \\ 2\pi - \theta & G < B \end{cases}, \text{ 其中 } \theta = \cos^{-1} \left[ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right] \quad (3-5)$$

从 RGB 颜色空间到 HSV 颜色空间的转换，还存在一种快速近似的转换公式 (Huang J. 1998, 孙兴华 2001) 如下：

$$V = \max(R, G, B) \quad (3-6)$$

$$S = \frac{V - \min(R, G, B)}{V} \quad (3-7)$$

令  $r' = \frac{V-R}{V-\min(R, G, B)}$ ,  $g' = \frac{V-G}{V-\min(R, G, B)}$ ,  $b' = \frac{V-B}{V-\min(R, G, B)}$ ，则有

$$H' = \begin{cases} 5 + b', & R = \max(R, G, B) \text{ 且 } G = \min(R, G, B) \\ 1 - g', & R = \max(R, G, B) \text{ 且 } G \neq \min(R, G, B) \\ 1 + r', & G = \max(R, G, B) \text{ 且 } B = \min(R, G, B) \\ 3 - b', & G = \max(R, G, B) \text{ 且 } B \neq \min(R, G, B) \\ 3 + g', & B = \max(R, G, B) \text{ 且 } R = \min(R, G, B) \\ 5 - r', & \text{其他} \end{cases} \quad (3-8)$$

$$H = 60 \times H' \quad (3-9)$$

由上述公式可知： $H \in [0, 360^\circ]$ ,  $S \in [0, 1]$ ,  $V \in [0, 1]$ 。

从 HSV 颜色空间到 RGB 颜色空间的转换公式如下 (孙兴华 2001, 章毓晋 1999)：



(1) 当  $H \in [0^\circ, 120^\circ)$  时

$$R = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos(H)}{\cos(60^\circ - H)} \right], \quad B = \frac{V}{\sqrt{3}} (1 - S), \quad G = \sqrt{3}V - R - B \quad (3-10)$$

(2) 当  $H \in [120^\circ, 240^\circ)$  时

$$G = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos(H - 120^\circ)}{\cos(180^\circ - H)} \right], \quad R = \frac{V}{\sqrt{3}} (1 - S), \quad B = \sqrt{3}V - G - R \quad (3-11)$$

(3) 当  $H \in [240^\circ, 360^\circ)$  时

$$B = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos(H - 240^\circ)}{\cos(300^\circ - H)} \right], \quad G = \frac{V}{\sqrt{3}} (1 - S), \quad R = \sqrt{3}V - G - B \quad (3-12)$$

### 3.2.4 CIEL\*a\*b\*和CIEL\*u\*v\*颜色空间

CIEL\*a\*b\*和 CIEL\*u\*v\*颜色空间是两种均匀的颜色模型。均匀颜色空间模型本质上仍是面向视觉感知的颜色模型，只是在视觉感知方面更为均匀。从视觉感知均匀的角度来说，人所感知到的两种颜色的距离，应该与这两种颜色在表达它们的颜色空间中的距离越成比例越好。换句话说，如果在某一颜色空间中，人所观察到的两种颜色的区别程度与该颜色空间中两点间的欧氏距离相对应，则称该空间为均匀的颜色空间（黄元元 2003）。

为了将易测的空间距离作为色彩感觉差别量的度量，1976 年，CIE（国际照明委员会）公布了两个标准的同一性空间 CIEL\*a\*b\*颜色空间和 CIEL\*u\*v\*颜色空间。这两个空间是在 CIE1931 年公布的 XYZ 空间的基础上得到的，下面我们给出了 RGB 颜色空间到 CIEL\*a\*b\*颜色空间和 CIEL\*u\*v\*颜色空间的转换公式（杨静宇等 1997）。

RGB→CIEXYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.2 \\ 0.177 & 0.812 & 0.011 \\ 0.0 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3-13)$$

CIEXYZ→CIEL\*a\*b\*

$$\begin{aligned} L^* &= 116(Y/Y_n)^{\frac{1}{3}} - 16 & \text{if } Y/Y_n > 0.008856 \\ L^* &= 903.3(Y/Y_n) & \text{if } Y/Y_n \leq 0.008856 \\ a^* &= 500 \left[ (X/X_n)^{\frac{1}{3}} - (Y/Y_n)^{\frac{1}{3}} \right] \\ b^* &= 200 \left[ (Y/Y_n)^{\frac{1}{3}} - (Z/Z_n)^{\frac{1}{3}} \right] \end{aligned} \quad (3-14)$$

CIEXYZ→CIEL\*u\*v\*

$$\begin{aligned} L^* &= 116(Y/Y_n)^{\frac{1}{3}} - 16 & \text{if } Y/Y_n > 0.008856 \\ L^* &= 903.3(Y/Y_n) & \text{if } Y/Y_n \leq 0.008856 \\ u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \end{aligned} \quad (3-15)$$

式中， $u' = \frac{4X}{X + 15Y + 3Z}$ ， $v' = \frac{9Y}{X + 15Y + 3Z}$ ， $(X_n, Y_n, Z_n)$  表示参考白色对应的值， $u'_n$  和

$v'_n$  表示参考白色的变换值, 其定义与  $u'$  和  $v'$  的定义相同。

CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 和 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色空间均是基于对立色理论和参考白点的 (Kenneth.R. Castleman 1998), 其中 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色模型与设备无关, 适用于显示器显示和根据加色原理进行组合的场合, 该模型中比较强调对红色的表示, 即对红色的变化比较敏感, 但对蓝色表示的比较粗糙。CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 颜色模型亦与设备无关, 适用于接近自然光照明的场合, 该模型比较强调对绿色的表示 (对绿色比较敏感), 其次是红色和蓝色。

对 CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 颜色空间和 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色空间而言, 完全可以用欧氏距离来测量颜色之间的差别, 而且, 这两个颜色空间不仅是同一性空间, 并且人眼对这两个空间的亮度分量  $L^*$  敏感, 对  $a^*$ 、 $b^*$  分量和  $u^*$ 、 $v^*$  分量的变化不甚敏感 (杨静宇等 1997)。

### 3.2.5 YCrCb 颜色空间

YCrCb 颜色空间是一种用于数字图像的颜色标准。该模型中  $Y$  代表了光源的亮度, 色度信息组合在  $Cr$ 、 $Cb$  中, 其中,  $Cr$  代表了光源中的红色分量,  $Cb$  代表了光源中的蓝色分量 (林福宗 2002)。

亮度给出了颜色亮或暗的程度信息, 是人在观察光照时感知亮度变化的心理上的度量, 可通过特定照明中的强度成分的加权和来计算。在 RGB 光源中, 光源的绿色分量对亮度影响最大, 蓝色分量对亮度影响最小, 因此, 亮度公式一般可表示如下:

$$Y=0.299R+0.587G+0.114B \quad (3-16)$$

由于人眼对于亮度的敏感程度大于对于色度的敏感程度, 所以完全可以让相邻的像素使用同一个色度值, 而人眼的感觉不会起太大的变化, 从而通过损失色度信息来达到节省存储空间的目的, 因此 YCrCb 颜色模型适合于图像文件的压缩, 同时它也被多种图像文件格式所采用, 如 JPEG、MPEG 等国际标准。

RGB 颜色空间与 YCrCb 颜色空间的转换关系如下:

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5 & -0.4187 & -0.0813 \\ -0.1687 & -0.3313 & 0.5 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (3-17)$$

除了上述颜色空间外, MPEG-7 支持的颜色空间还有 HMMD (Hue, Max, Min, Difference), 这是 MPEG-7 新提出的一种颜色空间 (Manjunath B.S. et al 2001)。

## 3.3 颜色量化

颜色量化是图像工程中的一项基本而重要的技术, 它是进行图像分割和对象提取的基础。自然界具有无限丰富的颜色, 真彩色图像具有  $2^{24}$  种颜色, 而在某些情况下 (如图像检索、印染等), 对颜色的数目有一定的限制。如何选取有代表性的若干种颜色, 并把各种颜色都归并到这些代表色上, 就是颜色量化需要解决的问题。

### 3.3.1 颜色量化的定义

颜色量化可形式化表述 (Oleg V. 1995, 何海涛等 2003) 为: 设  $c_i$  是颜色空间中的一个三维向量 (HSV, RGB 等颜色空间),  $C = \{c_i, i=1, 2, \dots, N\}$  表示输入图像中颜色的集合 ( $N$  表示颜色的数目),  $\bar{C} = \{\bar{c}_j, j=1, 2, \dots, K\}$  ( $K \ll N$ ) 表示输出图像中颜色的集合。颜

色量化是一个映射过程

$$q: C \rightarrow \bar{C} \quad (3-18)$$

颜色量化遵循距离最近准则：输入图像中的任一颜色  $c$  将被映射到调色板  $\bar{C}$  中距离最近的颜色  $\bar{c}$ ，即

$$\bar{c} = q(c) : \|c - \bar{c}\| = \min_{j=1,2,\dots,K} \|c - \bar{c}_j\| \quad (3-19)$$

同时，在颜色集合  $C$  中得到  $K$  个聚类  $S_k, k=1,2,\dots,K$

$$S_k = \{c \in C \mid q(c) = \bar{c}_k\} \quad (3-20)$$

$\bar{c}_k$  为  $K$  个聚类的聚类中心，它们组成输出图像的调色板。

颜色量化是把非常丰富的颜色量化到较少的颜色上去，因此不可避免的存在偏差，是一个有损的过程。颜色量化能否取得理想的效果，关键在于能否解决输入图像的整体层次和局部细节之间的矛盾。一个好的颜色量化算法需要在这对矛盾中找到合适的平衡点。

### 3.3.2 常用的颜色量化方法

目前常用的颜色量化方法大体可以分为分割算法和聚类算法两大类 (Zhou B. et al 2004)。代表性的分割算法有频度序列法 (Scheunders P. 1996)、中值裂分法 (耿国华等 1998)、八叉树法 (Gerrautz M. et al 1998) 等，其基本思想是将图像中出现的频率最高的  $K$  种色彩作为调色板，然后将其余颜色按照最小距离准则映射到调色板中。此类方法重构图像的层次感较丰富，但会丢失出现频率小的色彩，因而无法保留细节，使局部模糊。聚类算法则先选择若干聚类中心，然后按某种准则对颜色进行迭代聚合，直到合适的分类为止。典型的有  $K$  均值聚类算法 (Arthur R. et al 1997, Kanungo T. et al 2002, 何海涛等 2003)、模糊  $c$  均值 (Fuzzy  $c$ -means, FCM) 聚类算法 (Lim Y.W. et al 1990, 高新波等 2004)、学习向量法 (Yook K.L. 2004, 陈宗海等 2006) 等。聚类算法为近似最优算法，但需迭代运算，计算量大，而且量化结果往往依赖于初始聚类中心的选取。另外，聚类算法容易将相近的色彩合并，而破坏色彩的层次感。

频度序列法 (Scheunders P. 1996) 的实现分为两步：第一步选择图像中使用频率最高 256 种颜色作为颜色模板；第二步把其余颜色按最小距离准则映射到颜色模板中。该方法的特点是根据不同的图像选择不同的颜色模板，即可变颜色模板。它是一种简单的颜色量化方法，该算法的优点是算法复杂度低、实现容易，缺点是颜色量化效果不够理想。

中值裂分法 (耿国华等 1998) 是以集合  $R$ 、 $G$ 、 $B$  中方差最大的那个分量的中值作为裂分标准，不断地把颜色空间裂分成越来越小的区域，直至最终得到  $N$  个集合，各集合中心点的颜色即为颜色模板的颜色。由于该算法采用最大方差准则作为裂分标准，故所得到的颜色模板既能最大程度地反映图像颜色的使用频率，又能保证图像中使用较少却非常特殊的颜色不被丢失。

文献 (SCHETHNI R. et al 2001) 给出了另外的颜色量化分类方法，如表 3.1 所示。

表 3.1 颜色量化方法 (SCHETHNI R. et al 2001)


Color Discretization				
静态方法				动态方法
主要颜色的重要位	颜色空间划分	颜色空间聚类	参考颜色	图像分割

(1) 黑色区域: 所有  $v < 0.15$  的颜色均归入黑色, 令  $h = 0, s = 0, v = 0$ ;

(2) 白色区域: 所有  $s < 0.1$ , 且  $v > 0.8$  的颜色归入白色, 令  $h = 0, s = 0, v = 1$ ;

(3) 彩色区域: 位于黑色区域和白色区域以外的颜色, 其  $h, s, v$  保持不变。

在上述描述中,  $h$  表示色调,  $s$  表示饱和度,  $v$  表示亮度。通过上述处理, HSV 颜色模型进一步符合人的视觉特性, 更有利于图像颜色特征的提取, 改进后的 HSV 颜色模型如图 3.3 所示。



(g) 改进后的 V 分量

56

结合上述颜色量化方法及 Xiuqi Li 等 (Xiuqi LI et al 2002) 对进行颜色量化的基础上, 文献 (孙君顶等 2006) 提出将图像的颜色量化为 13 个级别, 量化方法如表 3.2 所示。

表 3.2 颜色量化

颜色 ID	颜色类型	色调范围	饱和度范围	亮度范围
1	白色	任意	<0.1	$\geq 0.85$
2	黑色	任意	任意	<0.15
3	灰色	任意	<0.1	(0.15, 0.85)
4	红色	$[350^{\circ}, 25^{\circ}]$	$\geq 0.1$	$\geq 0.15$
5	红-黄	$[25^{\circ}, 45^{\circ}]$		
6	黄色	$[45^{\circ}, 65^{\circ}]$		
7	黄-绿	$[65^{\circ}, 85^{\circ}]$		
8	绝色	$[85^{\circ}, 160^{\circ}]$		
9	绿-蓝	$[160^{\circ}, 180^{\circ}]$		
10	蓝色	$[180^{\circ}, 270^{\circ}]$		
11	蓝-紫	$[270^{\circ}, 290^{\circ}]$		
12	紫色	$[290^{\circ}, 330^{\circ}]$		
13	紫-红	$[330^{\circ}, 350^{\circ}]$		

有关更多的颜色量化方法, 可参阅文献 (SCHETHNI R. et al 2001, Van den Broek E.L. et al 2008, Shyh W.T. et al 2007, Teng S. 2003, Lu G. et al 1999, Teng S. et al 2006, Zhou B. et al 2004 等)。

## 3.4 全局颜色特征

鉴于颜色特征在基于内容图像检索中的重要作用, 很多基于颜色特征的图像检索算法被提了出来。总体来看, 基于颜色特征的图像检索算法主要集中在两个方面: 基于全局颜色特征检索方法和基于空间颜色特征检索方法。本节将主要讨论全局颜色特征的描述方法 (孙君顶 2005, 赵珊 2007, 姚琼等 2003, SCHETHNI R. et al 2001)。

### 3.4.1 颜色直方图

给定一幅图像  $(f_{xy})_{M \times N}$ ,  $f_{xy}$  表示像素点  $(x, y)$  处的颜色值,  $M \times N$  表示图像的尺寸, 图像所包含的颜色集记为  $C$ , 则图像的颜色直方可表示如下:

$$h_c = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f_{ij} - c), \quad \forall c \in C \quad (3-21)$$

颜色直方图作为一种重要的基于颜色特征进行图像检索的方法, 具有特征提取和相似度计算简便, 并且随图像尺度、旋转等变化不敏感的特点 (Swain M.J. et al 1991)。

但颜色直方图也存在 3 个缺点 (孙君顶 2005):

(1) 颜色直方图描述的是图像颜色的统计特性, 丢失了颜色的空间分布信息, 因此, 对于

两个颜色直方图相似的图像来说, 如果颜色的空间分布差别很大, 图像的内容会很不相同;

(2) 存在特征维数过高的问题;

(3) 由于对图像颜色的量化处理, 可能将视觉不同的颜色量化到同一区间, 也可能将视觉相同的颜色量化到不同的区间, 因而造成误检现象。

### 3.4.2 改进的颜色直方图方法

针对上述问题, 近年来许多改进的算法被提了出来。文献 (Li J. et al 2000, 卢汉清 2000) 采用彩色图像的主要颜色来构造颜色直方图, 由于忽略了那些数值较小的颜色区间, 改进后的颜色直方图对图像噪声的敏感程度降低了, 从而增强了检索效果。文献 (Stricker M. et al 1995, 刘忠伟 1998) 提出采用累加直方图方法和局部累加直方图来提高直方图特征进行图像检索的效率。梁艳梅等 (梁艳梅等 2003) 提出了一种基于直方图模糊相关的彩色图像检索算法, 应用模糊集理论的  $\alpha$ -级关系来定义彩色直方图的匹配色彩峰, 得到色彩峰的高度匹配关系, 通过综合所有色彩峰的高度匹配得出彩色直方图的相关值, 从而有效的提高图像的检索效率。为解决颜色量化时对颜色刚性划分带来的缺陷, 王炜等 (王炜等 2003) 提出了构造模糊直方图的方法来改善直方图特征的检索效果。王宇生等 (王宇生等 2002) 对每幅图像单独进行颜色量化, 计算其颜色直方图并排序, 在图像检索时根据颜色在图像中出现的频率和图像中各种颜色的对比强度来确定图像之间的相似度, 从而取得了更好的检索效果。有关更多的颜色直方图改进算法及相关颜色特征提取算法可参阅 (<http://www.visionbib.com/bibliography/compute105.html>)。

### 3.4.3 颜色不变量

为了克服光照的影响, Swain 等 (Swain M.J. et al 1991) 建议采用 Novak 的监督颜色常量算法 (supervised color constancy) (Novak C.L. et al 1991), 即在环境中放置已知反射率的色板, 以作为光照变化的参考, 但该方法在实际中却不易实现。

Funt B.V. 及 Finlayson G.D. (Funt B.V. et al 1995) 认为光照度在一定的区域内可以看做常量。他们对颜色值的对数求导 (拉普拉斯或方向导数), 计算直方图, 得到的实际是相邻颜色边界的长度, 以颜色变化率为索引, 以此排除光照成分对颜色的影响。这种方法当图像库在均匀光照下比颜色直方图的方法略差一些, 但当图像库中的图像受到光照影响时要好一些。

Finlayson 在文献 (Finlayson G.D. et al 1998) 建议使用颜色角度进行索引, 并通过实验证明该方法比颜色直方图 (Swain M.J. et al 1991) 和颜色常量 (Funt B.V. et al 1995) 的方法结果要好一些。Healey 和 Slater (Healey G. et al 1994) 推导出受光照影响小的颜色分布矩, 利用有限维模型提出了从直方图分布求取光照不变量。Gevers 等人 (Gevers T. et al 2000) 从 Shafer 双色反射模型, 推导出了受光照影响小的颜色模型如下:

$$\begin{aligned} l_1 &= \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_2 &= \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_3 &= \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \end{aligned} \quad (3-22)$$

式中,  $l_1, l_2, l_3$  表示颜色的三个通道。

更多有关颜色不变量的文献可参阅文献 (Gevers, T. et al 2004, Gevers, T. 2001, Ji Y.K. et al 2001, Finlayson, G. et al 2005 等)。

### 3.4.4 图像主色

一般地, 人眼在比较 2 幅图像的颜色时, 善于抓住各自的主色。虽然一幅自然图像的色彩极其丰富, 但视觉系统能够忽略次要的细节, 抓住起重要作用的主色。所谓主色, 是指图像的主色块的颜色, 一般是图像中物体的表面色彩或背景色彩, 占有较大的面积。主色的数量一般在数种到数十种之间。图像主色在 CBIR 中得到广泛应用 (徐旭等 1999, Ravishankar et al 1999, 袁昕等 2000, 袁昕等 2002a, 袁昕等 2002b, Mojsilovic A. et al 2002, Wang S.R. et al 2003, 孙君顶等 2006, Krishnan et al 2007)。Yang 等人 (Yang N.-C. et al 2008) 提出了一种 MPEG-7 的主色提取方法, 该方法没有采用复杂的聚类等方法来提取图像主色。其基本思想是首先将 RGB 空间进行粗略划分为不同的区间, 如图 3.5 所示, 然后采用每一个区间的中心作为该区间的量化结果。文中同时给出了主色间的相似度度量方法。图 3.6 给出了文中方法的一个示例。

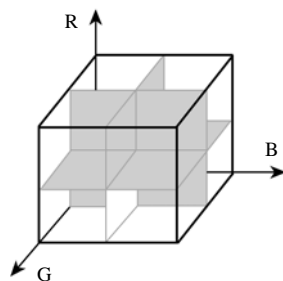


图 3.5 RGB 空间粗略划分







Query image Q	Target image F1	Target image F2
{(33,31,33),0.794240}	{(66,41,29),0.108795}	{(60,55,53),0.378306}
{(184,179,180),0.20576}	{(203,47,71),0.334035}	{(139,123,115),0.073598}
	{(207,193,59),0.067861}	{(198,194,188),0.548096}
	{(228,98,161),0.219045}	
	{(230,162,203),0.270264}	
		
		

图 3.6 图像主色及主色所占比例 (原始图像及采用主色的量化结果)

### 3.4.5 色调直方图

鉴于图像的颜色信息主要通过色调反映, 文献 (孙君顶等 2004a, 孙君顶等 2004b, 孙君顶等 2005) 分别采用色调直方图及主色调直方图来描述图像的颜色特征。

图 3.7 所示的是色调 H 的分布图 (黄元元 2003)。色调 H 是由颜色名称来辨别的, 它用角度  $0^\circ \sim 360^\circ$  来度量。接近  $0^\circ$  或  $360^\circ$  的色调是红色,  $120^\circ$  附近的色调是绿色, 而蓝色调是在  $240^\circ$  附近。在颜色轮上, 主要颜色沿一圆均匀分布, 其他一些次要的颜色位于主颜色之间。

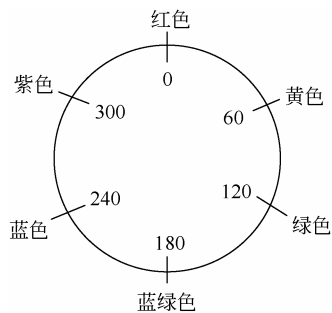


图 3.7 色调 H 示意图

为了获取一个独立于观察点的图像颜色描述符，需要一种颜色特征，在事先考虑到阴影、遮蔽以及亮度变化等因素的影响下，它仍能独立于对象表面的形状和观察的角度。人们已经发现，人眼对于颜色的色调尤为敏感，并且颜色的色调是一个独立于观察点的颜色特征（Gevers T. et al 1996, Gevers T. et al 1999）。

文献（孙君顶等 2004a）将色调均匀量化成 18 个不重叠的局部区间  $[20k, 20(k+1)]$ ,  $k = 0, 1, \dots, 17$ ，然后采用与颜色直方图相同的方法计算出此时的色调直方图，表示如下：

$$H_k = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(h_{ij} - k) \quad (3-23)$$

式中， $\delta(h_{ij} - k) = \begin{cases} 1, & h_{ij} = k \\ 0, & h_{ij} \neq k \end{cases}$ ， $M$ 、 $N$  表示图像的尺寸， $h_{ij}$  表示  $(i, j)$  处像素的色调值。

由于色调在色调轴上的分布实际上是连续过渡的，且各色调区间并不存在截然的界限，因此这样的划分结果就忽视了颜色在各个区间分界处的相似连续性，会造成与颜色量化时区间过大的相同问题（Zhang Y. J. et al 1997）。针对这种情况，我们对上述色调量化的方法又做了进一步的改进，即在第一次划分的基础上，再次改变色调区间的划分为  $[20k + 10, (20(k+1) + 10) \bmod 360]$ ,  $k = 0, 1, \dots, 17$ ，同时计算出此时的色调直方图，最后将这两个直方图逐项相加取平均作为图像的色调直方图。

当图像中的颜色并不能取遍所有可能取值时，统计直方图中会出现一些零值，这些零值的出现会对特征间相似度度量的计算带来影响，从而使得相似度度量并不能正确反映图像之间的颜色差别。为解决这个问题，可采用累加直方图方法（黄元元 2003）。累加直方图也是一个一维离散函数

$$\hat{H}(k) = \sum_{k'=0}^k \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(h_{ij} - k'), \quad k \in [0, 17] \quad (3-24)$$

累加直方图也可通过原直方图得出

$$\hat{H}(k) = \sum_{k'=0}^k H(k'), \quad k \in [0, 17] \quad (3-25)$$

累加直方图大大减少了原统计直方图中出现的零值，两种颜色在特征轴上的距离保持与它们之间的相似程度成正比。

### 3.4.6 颜色矢量角直方图

在 RGB 空间中度量两种颜色之间的色差有很多种方法，最常用的方法是采用欧氏距离

$$D(V_1, V_2) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \quad (3-26)$$

式中， $V_1 = [R_1, G_1, B_1]$ ,  $V_2 = [R_2, G_2, B_2]$ 。

欧氏距离的优点是简便易行，有清晰的物理意义，具有空间旋转不变性的特点。但是，通过我们前面对颜色空间的分析来看，RGB 颜色空间是不均匀的颜色空间，两种颜



色之间欧氏距离与两种颜色的视觉差异并不一致。如图 3.8 所示, 在 RGB 颜色空间中, 从人们的视觉感观来说, 颜色 C1 和 C2 的相似度要比 C3 和 C4 更大些, 然而它们两对颜色的欧氏距离却是相等的。

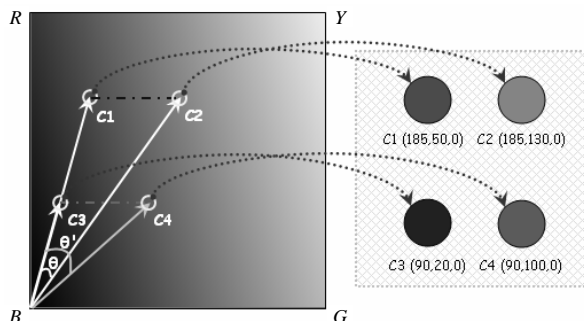


图 3.8 RGB 颜色空间中欧氏距离相等的两个颜色对 (C1, C2) 和 (C3, C4) 的视觉差异

为了度量不同颜色之间的差异, Dony 等人 (Dony R.D. et al 1999) 提出颜色矢量角概念, 其定义如下:

$$\sin(\theta)_{v_1, v_2} = \left( 1 - \frac{(v_1^T v_2)^2}{v_1^T v_1 v_2^T v_2} \right)^{\frac{1}{2}} \quad (3-27)$$

式中,  $v_1 = [R_1, G_1, B_1]^T$ ,  $v_2 = [R_2, G_2, B_2]^T$ 。

从图 3.8 我们还可以看出, 颜色对 (C1, C2) 之间的颜色矢量角要比 (C3, C4) 之间的颜色矢量角小, 这是由于颜色矢量角对亮度变化不敏感的原因, 因此利用颜色矢量角可以很好区分不同颜色在视觉上的差异。

在此基础上, 赵珊等人 (赵珊等 2006, 赵珊等 2007) 构造了颜色矢量角直方图来描述图像的颜色特征。利用颜色矢量角对色度和饱和度的敏感性来提取图像的边缘, 对于采用颜色矢量角提取的每个边缘像素点都具有一个最大矢量角, 这个特征反映了图像边缘取向的分布特性, 不同的图像内容对应不同的边缘分布, 具有不同的边缘取向分布特性。由于颜色矢量角具有对图像的亮度变化不敏感, 对色调和饱和度的变化敏感的特性, 能够很好的体现两种颜色间的视觉差异, 在彩色图像的边缘检测方面也得到很广泛的应用。为了利用颜色矢量角来检测图像的边缘信息, 首先需要建立一个  $3 \times 3$  的模板, 然后将这个模板按从左到右, 从上到下的顺序滑过图像中的每个像素, 计算每个像素和其 8 个邻域像素之间的 8 个颜色矢量角, 并找出最大的矢量角作为这个像素的矢量角。当一个像素对应的最大矢量角大于一个设定的阈值  $T$  时, 这个像素将被看做为边缘点, 反之为平滑点。如图 3.9 所示, 其中图 (a) 为示例图像, 图 (b)、(c)、(d)、(e) 分别为不同阈值下提取的边缘图像。

在此前提下, 构造颜色矢量角直方图作为表征图像内容的特征量。图像中的边缘按照强度可以分为两种, 即强边缘 (图像中大的景物、目标的边界) 和弱边缘 (由于噪声和精细纹理而得到的边缘)。其中强边缘包含了图像的主要信息, 是人们检索的依据; 而大量的弱边缘会使得到的直方图趋于平坦, 从而降低检索的性能。为了突出图像中的强边缘, 尽量避免弱边缘的影响, 引入一个控制因子来控制参与颜色矢量角直方图计算的边缘像素个数。由于颜色矢量角特有的对图像的亮度变化不敏感, 对色调和饱和度的变化敏感的特性, 因此, 构造的颜色矢量角直方图对光照变化、几何变化具有很强的鲁棒性。图 3.10

给出了三幅示例图像的颜色矢量角直方图。图 3.11 给出了不同光照情况下，颜色矢量角直方图的变化。

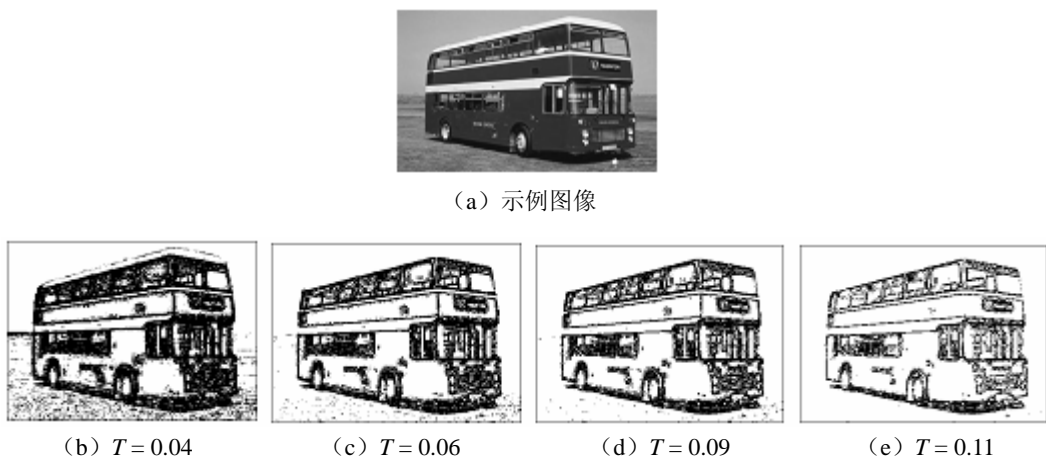


图 3.9 不同阈值下提取的边缘图像

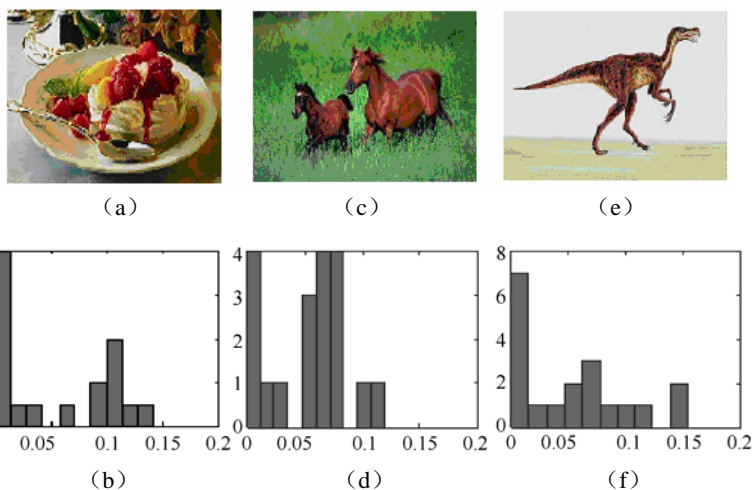


图 3.10 示例图像及其矢量角直方图

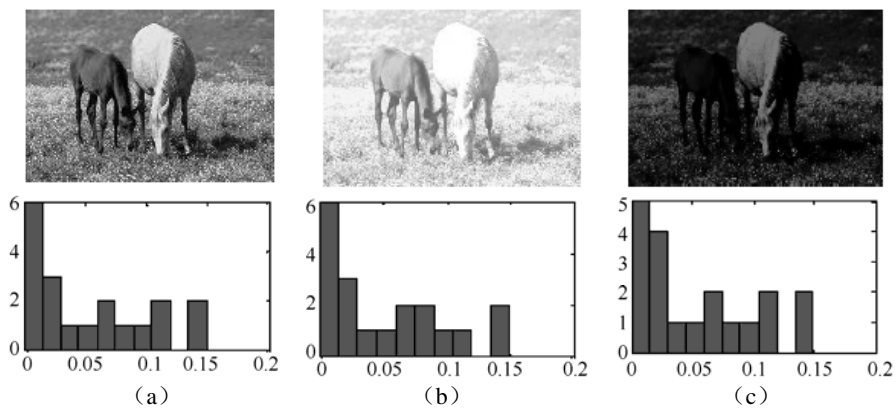


图 3.11 不同光照下的三幅图像及其颜色矢量角直方图

### 3.4.7 颜色矩

另一种非常简单且有效的颜色特征描述方法可以使用 Stricker 和 Orengo 所提出的颜色矩 (Color Moments) (Stricker M. et al 1995)。这种方法的数学基础在于图像中任何颜色分布均可以用它的矩来表示。此外, 由于颜色分布信息主要集中在低阶矩中, 仅采用颜色直方图特征的一阶矩 (Mean)、二阶中心矩 (Variance) 和三阶中心矩 (Skewness) 就足以表达图像的颜色特征。与颜色直方图相比, 该方法的另一个好处在于无需对颜色进行量化, 同时又降低了颜色特征的维数。颜色三个低阶矩的数学表达式为:

$$\mu_i = \frac{1}{n} \sum_{j=1}^n h_{ij} \quad (3-28)$$

$$\sigma_i = \left( \frac{1}{n} \sum_{j=1}^n (h_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (3-29)$$

$$s_i = \left( \frac{1}{n} \sum_{j=1}^n (h_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (3-30)$$

式中,  $h_{ij}$  表示第  $i$  个颜色通道分量中灰度为  $j$  的像素出现的概率,  $n$  表示灰度级数。因此, 图像的颜色矩一共有 9 个分量 (3 个颜色分量, 每个颜色分量有 3 个低阶矩), 颜色矩同其他颜色特征相比是非常简洁的, 但实验发现图像低阶矩的检索效率比颜色直方图的检索效率要低。在实际应用中为了避免低阶矩较弱的分辨能力, 往往将颜色矩同其他图像特征联合应用, 在利用其他特征进行图像检索前, 可首先采用颜色矩过滤缩小检索范围。

### 3.4.8 颜色熵

根据颜色直方图特性和信息论中信息熵的概念, John (John Z.M. 2000) 提出采用图像颜色的信息熵来表示图像的颜色特征, 从而将图像的颜色直方图由多维降低到一维, 然后采用  $L_1$  距离进行图像间的距离度量。

设图像的归一化颜色直方图表示为  $(h_1, h_2, \dots, h_n)$  如果我们将图像的颜色直方图看做是图像中不同颜色的像素在图像空间中出现的概率密度函数, 则根据信息熵理论, 图像的信息熵可表示为

$$E = - \sum_{i=1}^n h_i \log_2(h_i) \quad (3-31)$$

实验中发现, 虽然采用熵的方法有效地降低了图像直方图特征的维数, 但在利用颜色熵进行图像检索时, 其分辨能力是较低的。因此, 颜色熵特征往往也需要和其他图像特征相结合进行检索, 在利用其他图像特征进行图像检索前, 首先要利用图像信息熵缩小检索范围。

### 3.4.9 改进的颜色熵及颜色矩

John (John Z.M. 2000) 提出采用信息熵描述图像的颜色特征。将图像颜色直方图特征由多维降低到一维, 这能有效的克服直方图维数过高问题, 但是在采用图像信息熵表示图像的颜色特征时, John 并没有考虑熵的数学特性对这一特殊应用的影响。本节, 我们将分析熵的数学特性对图像检索效果的影响, 并讨论消除这些影响的方法。同时, 对颜色矩来说, 也存在类似的问题, 颜色矩表现出的也是一种统计特征, 视觉不同的颜色直方图如果

具有相似的概率分布，也将具有相似的颜色矩，本节也将讨论如何消除该影响。

### 1. 熵的基本知识

当我们收到一封信或是一封电报时，便能获得一定的信息，例如，某甲发了一封电报，内容是“平安到达北京”，那么对某乙来说，收到电报后消除了一些原先不能肯定的情况，这种原先不能肯定的情况称为先验不肯定性。假定电报内容只有“达到北京”几个字，对某乙来说还是不能肯定是否平安，只是部分的消除了不肯定性。对这种先验不肯定性，在信息论中用一个专有名词“信息”来表示（孟庆生 1986，金振玉 1991），也就是说，信息蕴涵于不肯定性之中，而不肯定性在概率论中是用随机事件或随机变量来描述的。

假设有一个有限事件集  $A = \{a_1, a_2, \dots, a_n\}$ ，其中各事件的出现相互独立，而且出现的概率分布为  $P = \{p_1, p_2, \dots, p_n\}$ ，把它们写在一起便称为概率空间（孟庆生 1986），即

$$X = \left\{ \begin{matrix} a_1, a_2, \dots, a_n \\ p_1, p_2, \dots, p_n \end{matrix} \right\} \quad (3-32)$$

式中， $p_i \geq 0$ ，且  $\sum_{i=1}^n p_i = 1$ 。事件集中的某一事件  $a_i$  的信息量用  $I_i$  表示为

$$I_i = \log_d \frac{1}{p_i}, 1 \leq i \leq n, d > 0 \quad (3-33)$$

式中，对数  $d$  表示事件  $A$  以  $d$  进制编码。 $I_i$  称为事件  $a_i$  的自信息量（Self information）。用  $I_i$  能表达事件出现的信息量是因为：

（1）当  $p_i = 1$  时， $I_i = 0$ 。 $p_i = 1$  表明事件  $a_i$  为必然事件，而必然事件没有不肯定性，因此其信息量为零，这与直观的看法相符。

（2） $I_i$  是事件  $a_i$  发生概率  $p_i$  的单调递减函数，即随事件  $a_i$  发生概率  $p_i$  的增加，信息量减小。

（3）用对数表示信息量，表明信息具有可加性，因为直观看法认为总信息量应为各信息量之和。

上述  $I_i$  表示事件集中某一事件的信息量。香农（Shannon）将整个事件集的平均信息量定义为熵（Shannon, C. E. 1948，金振玉 1991），即

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i \quad (3-34)$$

如果  $p_i = 0$ ，表示不可能事件，或者说事件根本没有发生，也就谈不上信息可言，因此规定  $0 \log_2^0 = 0$ 。

### 2. 信息熵的特性

#### （1）对称性

熵函数的所有变元可以互换，而不影响函数值。也就是说，不管  $p_i$  位置如何变化，它的值保持不变。即

$$H_n(p_1, p_2, \dots, p_n) = H_n(p_2, p_1, \dots, p_n) = H_n(p_{k(1)}, p_{k(2)}, \dots, p_{k(n)}) \quad (3-35)$$

式中， $p_{k(i)}$  表示第  $i$  个位置放置第  $k$  个元素。这个性质是明显的，因为  $H(X)$  是概率向量  $P$  的函数，所以只与随机变量的总体结构有关。

## (2) 非负性

由于  $0 \leq p_i \leq 1 (1 \leq i \leq n)$ , 则  $p_i \log p_i \leq 0$ , 从而有

$$H_n(\mathbf{P}) = H_n(p_1, p_2, \dots, p_n) \geq 0 \quad (3-36)$$

## (3) 极值性

$$H_n(p_1, p_2, \dots, p_n) \leq H_n\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = \log n \quad (3-37)$$

式 (3-37) 表明, 在离散的情况下, 集合  $\mathbf{X}$  中各事件等概率出现, 即服从均匀分布时, 熵值最大。

## (4) 扩展性

令  $\lim_{\varepsilon \rightarrow 0} \varepsilon \log \varepsilon = 0$ , 则

$$\lim_{\varepsilon \rightarrow 0} H_n(p_1, p_2, \dots, p_n - \varepsilon, \varepsilon) = H_K(p_1, p_2, \dots, p_n) \quad (3-38)$$

其含义是: 若集合  $\mathbf{X}$  有  $n$  个事件, 而集合  $\mathbf{X}'$  有  $n+1$  个事件, 但它和  $\mathbf{X}$  的差别只是多了一个概率近于零的事件时, 由熵函数的总体性可知, 概率小可以忽略不计。虽然概率小有较大的不肯定性, 但在熵的计算中取的是平均值, 因此它对于集合的熵值的贡献就可以忽略不计。

## (5) 确定性

若  $p_i = 1, p_l = 0$ , 且  $1 \leq i \leq n, 1 \leq l \leq n, i \neq l$ , 则

$$H(p_1, p_2, \dots, p_n) = 0 \quad (3-39)$$

即集合  $\mathbf{X}$  中, 只要有一个事件是必然事件, 则其余事件必为不可能事件。此时每个事件对熵的贡献都为零, 因此熵必为零。在这种情况下, 集合  $\mathbf{X}$  中的事件出现是完全确定的。

## 3. 改进算法

在利用熵进行图像检索时, 对检索效果产生重要影响的是熵的对称性问题。熵的对称性可表述为: 向量各分量的次序任意改变时, 熵值不变, 熵函数的取值只与向量的概率分布有关。因此, 对于具有不同视觉特征的直方图, 如果它们具有相同的概率分布, 则它们具有相同的信息熵。对于颜色矩来说, 也存在相应的问题。

如图 3.12 所示的三个直方图  $H_1$ 、 $H_2$  及  $H_3$ , 显然  $H_1$  和  $H_2$  具有一定的视觉相似性, 它们与  $H_3$  的视觉特性完全不同。但由于这三个直方图具有相同的概率分布特性, 因此它们具有相同的颜色熵和 (矩)。为了消除这种对称性对图像检索结果造成的影响以及考虑到人的视觉特征, 文献 (孙君顶等 2005, Sun J.D. 2006a, Sun J.D. 2006b) 提出了直方图排序法、直方图面积法、直方图排序法和直方图面积法的线性组合来消除这种影响。

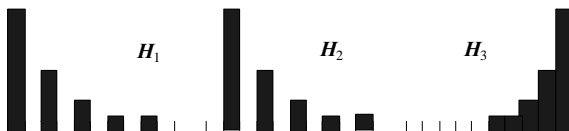


图 3.12 颜色直方图示例

#### 4. 直方图排序法

虽然图 3.10 所示三个直方图具有相同的概率分布特征,但很明显,不同概率在直方图中的位置完全不同,我们可借助于这种位置特征来进行区分。

设  $\mathbf{H}$  表示某幅图像的颜色直方图,我们采用某种排序方法(如冒泡排序)将直方图  $\mathbf{H}$  的各个分量由大到小(或从小到大)进行排序,假设排序后的颜色直方图为  $\mathbf{H}'$ ,在排序过程中直方图各分量移动的总次数记为  $M$ 。由颜色熵及矩的特性可知,直方图  $\mathbf{H}$  和  $\mathbf{H}'$  具有相同的信息熵和矩,所以这种排序不会造成直方图信息熵和矩的改变。对于具有相同或相近概率分布的直方图来说,如果直方图中各分量的次序不同,在对直方图进行排序后,各分量的移动次数也将不同,因此,采用排序时各分量移动次数的差异可有效地消除这种影响。

按照以上分析,在计算信息熵和矩时引入了加权函数

$$f_1(\mathbf{H}) = 1 + M_{\mathbf{H}} / M_{\max} \quad (3-40)$$

式中,  $\mathbf{H}$  表示任意的颜色直方图;  $M_{\mathbf{H}}$  表示直方图  $\mathbf{H}$  经过排序后,各分量移动的总次数;  $M_{\max}$  表示排序时,直方图分量需要移动的最大次数,即对逆序直方图进行排序时直方图各分量移动的次数。

从而图像信息熵和矩可分别表示为:

$$E(\mathbf{H}) = -f_1(\mathbf{H}) \sum_{i=1}^n h_i \log_2(h_i) \quad (3-41)$$

$$\begin{aligned} \mu &= f_1(\mathbf{H}) \frac{1}{N} \sum_{i=1}^N p_i \\ \sigma &= f_1(\mathbf{H}) \left( \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^2 \right)^{\frac{1}{2}} \\ s &= f_1(\mathbf{H}) \left( \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^3 \right)^{\frac{1}{3}} \end{aligned} \quad (3-42)$$

设  $E_0$  代表图 3.8 直方图的信息熵,  $\mu_0$ 、 $\sigma_0$  及  $s_0$  表示它们的颜色矩。如果按照从小到大的顺序将三个直方图进行排列,则  $M_{\mathbf{H}_1} = 25$ ,  $M_{\mathbf{H}_2} = 20$ ,  $M_{\mathbf{H}_3} = 0$ ,  $M_{\max} = 45$ ;如果我们采用  $L_1$  距离进行度量,对于直方图熵来说有

$$d_{L_1}(\mathbf{H}_1, \mathbf{H}_3) = 0.56E_0 > d_{L_1}(\mathbf{H}_2, \mathbf{H}_3) = 0.44E_0 > d_{L_1}(\mathbf{H}_1, \mathbf{H}_2) = 0.11E_0 \quad (3-43)$$

对于颜色矩来说有

$$\begin{aligned} d_{L_1}(\mathbf{H}_1, \mathbf{H}_3) &= 0.56(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\ &> d_{L_1}(\mathbf{H}_2, \mathbf{H}_3) = 0.44(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\ &> d_{L_1}(\mathbf{H}_1, \mathbf{H}_2) = 0.11(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \end{aligned} \quad (3-44)$$

式中,  $\omega_i (1 \leq i \leq 3)$  代表权值。

显然,从上述式子的计算结果可以看出,直方图  $\mathbf{H}_1$  和  $\mathbf{H}_2$  间的距离较小,而它们与直方图  $\mathbf{H}_3$  的距离较大,该结果也与人类的视觉特性一致。

#### 5. 直方图面积法

在采用直方图排序法进行改进时,仅仅考虑了直方图中各分量在整个向量中的位置因

素，并没有考虑该位置的分量所对应概率的大小，为此，我们进一步提出了直方图面积法。

设直方图  $H = \{p_1, p_2, \dots, p_n\}$ ，定义直方图的面积为

$$A_H = \sum_{i=1}^n (p_i \times i) \quad (3-45)$$

由式 (3-45) 可知，当  $p_n = 1$  时直方图的面积达到最大值  $A_{\max} = n$ 。对于具有相同或相近概率分布的直方图来说，如果直方图中各分量的次序不同，直方图所对应的面积一般也是不同的。因此，采用直方图面积也可有效地消除熵的对称性的影响。为此，在计算直方图的信息熵时，可引入了加权函数  $f_2(H)$ ，其定义如下：

$$f_2(H) = 1 + \frac{A_H}{A_{\max}} \quad (3-46)$$

从而图像信息熵和矩可分别表示为：

$$E(H) = -f_2(H) \sum_{i=1}^n h_i \log_2(h_i) \quad (3-47)$$

$$\begin{aligned} \mu &= f_2(H) \frac{1}{N} \sum_{i=1}^N p_i \\ \sigma &= f_2(H) \left( \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^2 \right)^{\frac{1}{2}} \\ s &= f_2(H) \left( \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^3 \right)^{\frac{1}{3}} \end{aligned} \quad (3-48)$$

在图 3.10 中，同样设  $E_0$  代表直方图的信息熵， $\mu_0$ 、 $\sigma_0$  及  $s_0$  表示它们的颜色矩，同时设  $H_1 = \{1/2, 0, 1/4, 0, 1/8, 0, 1/16, 0, 1/16, 0\}$ ， $H_2 = \{0, 1/2, 0, 1/4, 0, 1/8, 0, 1/16, 0, 1/16\}$ ， $H_3 = \{0, 0, 0, 0, 0, 1/16, 1/16, 1/8, 1/4, 1/2\}$ 。则有  $A_{\max} = 10$ ， $A_{H_1} = 2.88$ ， $A_{H_2} = 3.88$ ， $A_{H_3} = 9.06$ 。采用直方图面积法，直方图间信息熵间的距离分别为： $d_{L_1}(H_1, H_2) = 0.1E_0$ ， $d_{L_1}(H_1, H_3) = 0.62E_0$ ， $d_{L_1}(H_2, H_3) = 0.52E_0$ 。若采样颜色矩，则有

$$\begin{aligned} d_{L_1}(H_1, H_3) &= 0.62(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\ &> d_{L_1}(H_2, H_3) = 0.52(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\ &> d_{L_1}(H_1, H_2) = 0.10(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \end{aligned} \quad (3-49)$$

从计算结果可以看出，该结果也同人的视觉感知一致。同时，采用直方图面积法，直方图  $H_1$  及  $H_2$  间的距离变的更小了，而它们与直方图  $H_3$  的距离增大了，这一结果与人类的视觉特性更为接近，因此利用直方图面积法比直方图排序法能取得更好的效果。这是由于面积法既考虑了直方图中各分量的位置因素，又考虑了该位置的分量所对应的概率大小。

## 6. 直方图排序法与面积法的线性组合

虽然实验中发现采用直方图面积法的改进效果总体上优于直方图排序法，但对于一些特殊的情况，直方图面积法的改进效果反而比不上直方图排序法。例如，对于图 3.13 所示的两个直方图 (a) 和 (b) 来说，它们具有相同的信息熵，并且存在  $p_1m_1 + p_2m_2 = p_2m_3 + p_1m_4$ ，即具有相同的直方图面积，因此采用面积法是不能区分出直方图 (a) 和 (b) 的，但是如果采用直方图排序法，则可有效的区分出直方图 (a) 和 (b)。

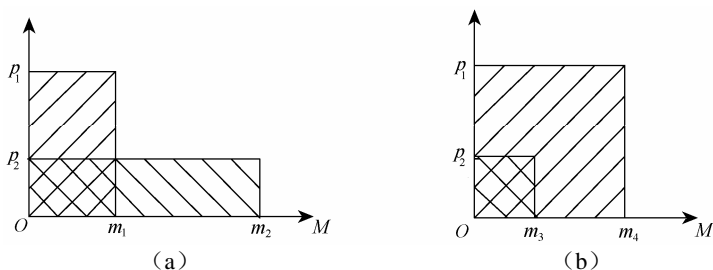


图 3.13 特殊情况的直方图

为此，我们可以进一步采用直方图排序法和直方图面积方法的线性组合来进一步增强图像的检索效果。采用两种方法线性组合的权函数可表示为

$$f_3(\mathbf{H}) = \gamma_1 f_1(\mathbf{H}) + \gamma_2 f_2(\mathbf{H}) \quad (3-50)$$

式中， $\gamma_1$ 、 $\gamma_2$  为权值系数，且  $\gamma_1 \in [0,1]$ ， $\gamma_2 \in [0,1]$ ， $\gamma_1 + \gamma_2 = 1$ 。

由于直方图排序法仅仅考虑了直方图中不同分量的位置关系，没考虑相应位置所对应概率的大小，而面积法既考虑了直方图中各个概率的位置关系，又考虑了该位置处所对应概率的大小，所以面积法在一定程度上优于直方图排序法。从稍后的实验中我们也可以看出这一点。因此，对于式 (3-50) 在选择权值时，往往取  $\gamma_1 > \gamma_2$ 。

### 7. 三种改进算法检索性能对比

本实验仅仅是为了测试直方图排序法、直方图面积法以及这两种方法线性组合的检索性能。在图像库 (<http://wang.ist.psu.edu/~jwang/test1.tar>) 中选取 10 个类别共 150 幅图像组成测试库，选取 HSV 颜色空间作为处理彩色图像的颜色空间，并将色度均匀量化为 18 个级别，饱和度和亮度分别量化为 3 个级别，将图像的颜色量化为 162 柄。然后分别采用未改进方法、直方图排序法、直方图面积法和直方图排序法及直方图面积法的线性组合四种方法来提取图像的信息熵。最后采用  $L_1$  距离来进行距离度量，对于式 (3-50) 取  $\gamma_1 = 0.2$ ， $\gamma_2 = 0.8$ 。实验中在测试库中任意选取 5 个类别的图像，并从每一类图像中分别选取 5 幅图像作为例子图像，总共形成 25 次检索。分别计算这 25 次查询的“检索率”和“精确度”的平均值作为最终的结果。图 3.14 给出了本实验四种方法 25 次检索的平均

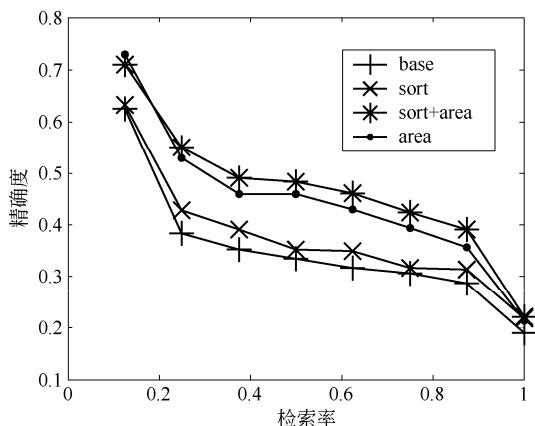


图 3.14 四种算法的“检索率”与“精确度”对比曲线

“检索率”与“准确度”对比曲线。其中，sort+area 代表利用直方图排序法及直方图面积法的线性组合进行检索的结果；area 代表采用直方图面积法进行检索的结果；sort 代表利用直方图排序法进行检索的结果；base 代表未采用任何改进算法的检索结果。从该图可以看出，采用直方图面积法比直方图排序法的检索效果要好，这是由于直方图面积法既考虑了图像直方图中各分量在整个向量中的位置因素，又考虑了相应位置处所对应分量的概率大小，而采用直方图排序法和直方图面积法的线性组合取



得了最好的检索效果。

### 3.5 空间颜色特征

上述颜色直方图、颜色矩和颜色熵等所描述的颜色特征是图像的全局颜色特征，不包括图像颜色的空间分布特征，因此仅仅利用这些特征进行图像检索时极易造成误检现象。如图 3.15 所示，四幅图像具有相同的直方图，但其视觉特征却完全不同，采样统计特征很难将它们区分开来。基于此，许多用于提取图像颜色空间分布特征的方法就被提了出来。



图 3.15 直方图相同，空间分布完全不同的图像 (SCHETHNI R. et al 2001)

#### 3.5.1 改进的颜色直方图法

为了在图像的颜色直方图中包含图像颜色的空间分布信息，图像的边缘信息被融入到图像的颜色直方图中 (Constantium N. et al 2000)。设  $\Delta(i, j)$  表示在像素  $(i, j)$  处拉普拉斯算子计算的结果，则加权的颜色直方图定义为：

$$\tilde{h}(c) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f_{ij} - c) \frac{1}{1 + \Delta(i, j)}, \quad \forall c \in C \quad (3-51)$$

$$\tilde{h}(c) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f_{ij} - c) \Delta(i, j), \quad \forall c \in C \quad (3-52)$$

其中，式 (3-51) 强调的是处于平坦区域的像素，由于处于边缘的像素， $\Delta(i, j)$  计算结果较大。与此相反，式 (3-52) 主要加强了边缘像素的影响，减弱了平坦区域像素的影响。

在式 (3-51) 和式 (3-52) 的基础上，Constantium 等 (Constantium N. 2000) 又提出了下述改进的加权颜色直方图

$$\tilde{h}(c) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f_{ij} - c) (k + \Delta(i, j))^\alpha, \quad \forall c \in C \quad (3-53)$$

式中，当  $\alpha = 0$  时，退化为一般的直方图；当  $\alpha < 0$  且  $k \neq 0$  时，属于连续区域内的像素得到了加强；当  $\alpha > 0$  且  $k \rightarrow 0$  时，属于图像边缘的像素得到了加强。

Guoping Qiu 等人 (Qiu G.P. et al 2003) 根据人的视觉特性与频域中频率的关系，提出采用高通滤波器、低通滤波器以及带通滤波器将图像的亮度分量转化为不同的频率分层，进一步将彩色图像转化为不同的频率分层，然后对于每一频率分层采用颜色直方图方法来进行索引，从而有效的提高图像的检索效果。宋擒豹等 (宋擒豹等 2006) 提出了一种颜色-位置直方图方法，该直方图在不失传统直方图稳健性的前提下，将图像的颜色信息和空间位置信息有机的融合起来，在反映颜色统计信息的同时也记录了颜色分段虚拟边界的位置信息，较好的解决了传统直方图存在的问题。有关更多的直方图的改进方法，可参阅 (<http://www.visionbib.com/bibliography/compute105.html>) 提供的文献。

### 3.5.2 颜色聚合向量

针对颜色直方图和颜色矩等无法描述图像颜色空间分布的信息, Pass 等人 (Pass G et al 1997) 提出了颜色聚合向量 (Color Coherence Vector) 的方法。其核心思想是, 将属于颜色直方图每一个区间内的像素分为两部分, 如果该区间内的某些像素所占据的连续区域的面积大于给定的阈值, 则将该区域内的像素作为聚合像素, 否则作为非聚合像素。假设  $\alpha_i$  和  $\beta_i$  分别代表直方图的第  $i$  个区间中聚合像素和非聚合像素的数目, 图像的颜色聚合向量可表示为

$$\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n) \rangle \quad (3-54)$$

而  $\langle \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n \rangle$  就是该图像的颜色直方图。由于颜色聚合向量包含了图像颜色的空间分布信息, 因而取得了相对于颜色直方图更好的检索效果。

### 3.5.3 颜色相关图

颜色相关图 (Color Correlogram) 是利用图像中像素间的颜色关系来描述图像颜色空间分布的另一种表达方式 (Huang J. 1998)。颜色相关图不但刻画了某一颜色的像素占整个图像的比例, 还反映了不同颜色对之间的空间相关性。对于任意的图像  $I$ , 假设  $I_{c(i)}$  表示图像中颜色为  $c(i)$  的所有像素的集合, 则颜色相关图可以表示为

$$r_{i,j}^{(k)} = \Pr_{p_1 \in I_{c(i)}, p_2 \in I} [p_2 \in I_{c(j)}, |p_1 - p_2| = k | p_1 \in I_{c(i)}] \quad (3-55)$$

式中  $i, j \in \{1, 2, \dots, n\}$ ,  $n$  表示图像颜色级数,  $k \in \{1, 2, \dots, d\}$ ,  $d$  表示在计算图像的相关图时所设定的像素间的最大距离,  $|p_1 - p_2|$  表示像素  $p_1$  和  $p_2$  间的距离。此时颜色直方图可以看做是一张用颜色对  $\langle i, j \rangle$  索引的表, 其中  $\langle i, j \rangle$  的第  $k$  个分量表示图像中颜色为  $c(i)$  和  $c(j)$  且距离为  $k$  的像素对出现的概率。如果考虑到任何颜色间的相关性, 颜色相关图就会变得非常复杂和庞大 (空间复杂度为  $O(n^2d)$ )。一种简化的方法是颜色自相关图 (Color auto-Correlogram), 它仅仅考虑图像空间中具有相同颜色像素间的空间关系, 因此空间复杂度降为  $O(nd)$ 。

### 3.5.4 局部颜色特征

目前从划分局部区域的角度来说, 常用的划分方法包括: 基于固定块的图像分割、基于手工的区域分割、采用交互半自动的区域分割以及一些自动的颜色分割方法。同时, 局部区域颜色信息主要采用平均颜色、主颜色、颜色直方图、颜色矩和二进制颜色集等来表示 (李向阳等 2001)。

颜色集是一种常用的局部颜色特征描述方法, 是对颜色直方图的一种近似 (Smith J.R. 1997)。在该方法中, 首先将图像从 RGB 颜色空间转化成视觉均衡的颜色空间 (如 HSV), 并将颜色空间量化成若干个柄; 然后, 用色彩自动分割技术将图像分为若干区域, 每个区域用量化颜色空间的某个颜色分量来索引, 从而将图像表达为一个二进制的颜色索引集。在图像匹配中, 需要比较不同图像颜色集之间的距离和色彩区域的空间关系 (包括区域的分离、包含和相交等操作, 每种对应于不同的评分), 因为颜色集表达为二进制的特征向量, 可经构造二分查找树来加快检索速度, 这对于大规模的图像集合十分有利。

Hsu 等 (Hsu W. et al 1995) 提出从图像中选择一些代表颜色, 然后将图像划分为矩形区域, 每个区域以一种主要的单一颜色作为代表, 两幅图像之间的相似性采用两个图像之间具有相似颜色区域的重叠程度来表示。文献 (Dimai A. 1997, Sciascio E.D. et al 1999, Yoo H.W. et al 2002) 利用固定尺寸的网格将图像划分为不同的区间, 并提取每一区间的颜色直方图作为图像的索引, 但采用该方法, 即使图像中不包含某些颜色, 局部颜色直方图特征中仍包含该颜色特征, 因而存储代价十分昂贵。为了改变这种情况, Stehling (Stehling R.O. et al 2002) 提出了一种改进的方法 CSH (Color-Shape Histograms), 它统计不同颜色在每一个局部分块的统计特征, 这样如果图像中不包含某种颜色, 就不必要进行存储。

通常, 一幅图像具有若干表示主题内容的主题画面, 由于对原图像事先缺乏任何先验知识, 无法确定这些主题画面在图像中的位置和大小。而尽可能将表现图像内容的主题画面划分到同一分块内是一种好的选择, 采用固定网格的划分方法是一种刚性的划分, 往往会将图像主题信息划分到不同的网格中。基于此, 文献 (何清法等 2001, Sebe N. 1999) 提出采用多分辨率的划分方法。采用不同分辨率的网格将图像划分为不同等级, 并且在进行图像划分时, 各分块间可采用重叠的方式, 提取每一分块的主颜色或颜色矩作为分块的颜色特征。

Stricker 和 Dimai (Stricker M. et al 1997) 认为图像中最有意义的区域是位于图像的中心, 因而将图像划分为 5 个模糊的区域, 其中中心区域为一个椭圆形, 同时将椭圆外的图像进行一定的模糊化, 以确保图像中的颜色信息对于图像小的平移和旋转影响不敏感, 然后采用每一区域的颜色矩作为图像的索引特征。图像区域的一些划分方法如图 3.16 所示。



图 3.16 几种图像区域划分方法 (SCHETHNI R. et al 2001)

Rao 等人 (Rao A. et al 1999) 将某种颜色在图像空间中的分布看做 2-D 平面, 并以该颜色的质心为中心, 将该 2-D 平面划分为不同的环形区间, 然后采用环形颜色直方图 (annular color histogram)、角度颜色直方图 (angle color histogram) 等作为图像索引。在此基础上, Cinque 等人 (Cinque L. et al 1999) 提出采用空间颜色直方图 (Spatial-Chromatic Histogram) 作为图像特征, 但空间颜色直方图不满足图像的尺度不变性, 为此, Lim 等人 (Lim S. et al 2003) 提出采用类似地理学中统计“区域人口分布”的方法对环形颜色直方图进行了归一化处理, 以满足所提取特征的尺度不变特性。

### 3.6 颜色空间分布熵

在文献 (孙君顶等 2006a, Junding Sun et al 2006b, 孙君顶等 2005) 中, 提出利用信

息熵来描述图像颜色的空间特征，并提出了相应的增强算法，具体介绍如下。

### 3.6.1 环形颜色直方图

Rao 等 (Rao A. 1999) 提出采用环形颜色直方图 (Annular Color Histogram) 来描述图像颜色的空间分布特征。设  $I$  表示任意一幅图像， $I(x, y)$  表示像素  $(x, y)$  处的颜色值，设  $A_i = \{(x, y) | (x, y) \in I, I(x, y) = i, 1 \leq i \leq n\}$  表示图像中颜色为  $i$  的所有像素的集合，其中  $n$  表示图像颜色的量化级数。设  $|A_i|$  表示集合  $A_i$  中像素的数目， $O_i = (x_i, y_i)$  为图像中颜色为  $i$  的所有像素的质心。 $x_i$  和  $y_i$  定义如下：

$$x_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} x; \quad y_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} y \quad (3-56)$$

设  $r_i$  表示图像中颜色为  $i$  的像素同其质心的最大距离，其定义如下：

$$r_i = \max_{(x,y) \in A_i} \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (3-57)$$

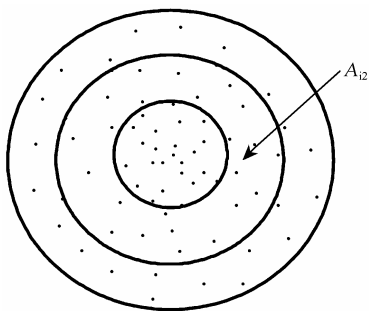


图 3.17 环形颜色直方图

对于给定的一个正整数  $N$ ，把  $r_i$  分为  $N$  等份，然后以质心  $O_i$  为圆心，以  $(j \times r_i)/N$  为半径 ( $1 \leq j \leq N$ ) 画圆可得到  $N$  个环。那么由内至外每个环和  $A_i$  的交点把  $A_i$  分为  $A_{i1}, A_{i2}, \dots, A_{iN}$ ，我们称  $A_{i1}, A_{i2}, \dots, A_{iN}$  为  $A_i$  的一个分割，如图 3.17 所示。设  $|A_{ij}|$  表示环形区间  $j$  内颜色为  $i$  的像素数目，则  $(|A_{i1}|, |A_{i2}|, \dots, |A_{iN}|)$  就构成了颜色  $i$  的环形分布直方图。由于质心  $O_i$  具有平移和旋转不变性，所以求取的环形颜色直方图同样具有平移和旋转不变性。

### 3.6.2 改进的环形颜色直方图

针对 Rao 等提出的环形颜色直方图，本节给出了改进方法。

(1) 采用  $L_\infty$  距离定义像素间的空间距离

在 (Rao A. 1999) 中，两像素间的空间距离采用  $L_2$  距离来进行度量，因而计算复杂度较高，这里采用  $L_\infty$  距离来定义图像像素间的距离。设像素  $p_1 = (x_1, y_1) \in I, p_2 = (x_2, y_2) \in I$ ，像素间的空间距离可定义为

$$|p_1 - p_2| = \max\{|x_1 - x_2|, |y_1 - y_2|\} \quad (3-58)$$

采用该方法，图像被划分为不同的矩形区间。

(2) 尺度不变性

(Rao A., 1999) 提出的环形颜色直方图虽然满足旋转不变性和平移不变性，但是它与图像中某一环形区间内相同颜色像素的数目相关，因此，该环形颜色直方图受图像的尺寸影响较大，不具有尺度不变特性。为了使环形颜色直方图满足尺度不变性，我们采用如下的方法对环形颜色直方图进行了归一化处理

$$p_{ij} = |A_{ij}| / |A_i| \quad (3-59)$$

这样，经过归一化后的环形颜色直方图满足旋转、平移和尺度不变性。

### 3.6.3 空间分布熵

通过上述处理,图像中的每一种颜色均对应一个环形分布直方图,若采用 Rao (Rao A. 1999) 的方法,则直接采用环形颜色直方图结合图像的颜色直方图来进行图像检索。但由于每一种颜色均对应一环形颜色直方图,因此这将大大增加存储这些特征所需的存储空间,同时还会造成检索速度的下降。为此我们利用熵的特性,提出采用颜色空间分布熵来描述颜色的空间分布特征。

根据式 (3-59) 和熵的定义,颜色  $i$  的空间分布熵表示如下:

$$e_i = -\sum_{j=1}^N p_{ij} \log_2(p_{ij}) \quad (3-60)$$

空间分布熵反映了具有某种颜色的像素在图像空间中的平均分散程度。颜色空间分布熵越大,表明具有该颜色的像素在图像空间中越分散;否则,表明具有该颜色的像素在图像空间的分布比较集中。采用颜色空间分布熵可有效的表征颜色的空间分布特征,颜色特征的维数也将大大降低。同时,由于归一化后的环形颜色直方图满足平移、旋转和尺度不变性,因此,颜色空间分布熵也满足平移、旋转和尺度不变性。

### 3.6.4 加权颜色空间分布熵

在人眼看来,对于具有相同颜色的像素在空间的分布来说,若该颜色的像素在空间中的分布越集中,则该种颜色对人的视觉刺激越大;相反,若该颜色的像素在空间中的分布越分散,则对人眼的视觉影响越小。对于空间分布熵来说,熵越大表明具有某颜色的像素在图像空间的分布越分散,因此对人眼的视觉刺激越小;熵越小表明具有某颜色的像素在图像空间的分布越集中,因此对人眼的视觉刺激越大。

一般来说,距离质心越近的环形区间,其像素分布越集中,距离质心越远的环形区间,其像素分布相对分散。为此,在计算颜色分布熵时,引进权函数  $f_4(j)$  ( $j$  表示不同的环形区间) 来反映不同环形区间对图像内容(人眼视觉的刺激程度)的贡献程度。结合人类的视觉特征及熵的特性,权函数的设置满足:距离质心越近的区间应赋予较小的权值,距离质心较远的区间应赋予较大的权值。从而权函数定义如下:

$$f_4(j) = 1 + \frac{j}{N} \quad (3-61)$$

而图像的加权空间分布熵表示为

$$e_i = -\sum_{j=1}^N f_4(j) p_{ij} \log_2(p_{ij}) \quad (3-62)$$

由于不同的环形区间引入不同的权值,这样就消除了由于空间分布熵相近使得空间分布直方图不同而对检索结果带来的影响,该方法也可以有效的解决熵的对称性对图像检索结果所带来的影响。

### 3.6.5 消除孤立分布小颜色块的影响

就人类的视觉特征来说,人往往对图像中大的颜色块比较敏感,而对于图像中孤立分布的颜色则不敏感,因此我们可以借助于图像颜色的空间分布熵来消除这种因素对图像检索结果的影响。

设  $p_i$  和  $e_i$  分别为颜色  $i$  的直方图特征和空间分布特征。如果  $p_i < \alpha_1$ ，并且  $e_i > \alpha_2$ （ $\alpha_1$  和  $\alpha_2$  为预先设定的两个阈值），我们将颜色  $i$  定义为图像中孤立分布的小颜色块。也就是说如果颜色为  $i$  的像素在图像中出现的概率很小，小于给定的阈值  $\alpha_1$ ，同时颜色为  $i$  的像素在图像空间中却很分散，其分布熵大于给定的阈值  $\alpha_2$ ，在进行图像检索时我们可不考虑颜色  $i$  对图像检索的影响。

## 3.7 位平面熵

由 3.3 节图像信息熵的定义可知，图像的信息熵体现了图像全局的统计特性——只与图像中各灰度出现的概率有关，具有相同灰度概率分布的两幅图像具有相同的信息熵。因此，图像的信息熵同图像的全局直方图一样，仅仅考虑了图像的全局统计信息，丢弃了图像的空间分布信息，而具有相同信息熵的图像可能在视觉上是完全不同的，所以图像的全局信息熵不足以反映出图像间的差异。为此，本节引入位平面熵的概念，使得图像特征能用一个熵矢量来描述，同时，对于熵相同的两幅图像所提取的位平面熵却不同，从而包含了图像像素的空间分布信息，解决了熵相同而图像内容不同的问题（赵珊等 2007）。

### 3.7.1 位平面分解与位平面

位平面编码属于无损图像压缩技术，是一种基于将多灰度图像分解为一系列二值图，然后对每一幅二值图再用二元压缩方法进行压缩的技术。这种技术不但能消除或减少编码冗余，也能消除或减少图像中像素间的冗余，其中的关键技术涉及位平面分解和编码方法。

数字图像可分为二值（黑和白）和多值图像两大类。对于多值图像来说，可以利用位平面分解的方法将其二值化。所谓位平面的含义就是将一个像素的灰度值分解为二进制值，所有同权值的位（0 或 1）构成的平面叫位平面。例如，一个灰度为 256 的图像，每个像素占一个字节，即 8 个二进制位，按从高位到低位的排列为  $b_7b_6b_5b_4b_3b_2b_1b_0$ ，那么所有像素的  $b_0$  位就构成第 0 个位平面， $b_1$  位就构成第 1 个位平面，以此类推，将图像分解为 8 个位平面。这样，一幅灰度图像就可以被看做是 8 个位平面的叠加，每个位平面被看做是一幅二值图像。

设  $I$  表示一幅灰度图像， $I(i, j)$  为其中的一个像素，则对该像素的位平面分解可定义为

$$g_t(i, j) = B_t(I(i, j)) = \begin{cases} 1 & (\text{Int}(\frac{I(i, j)}{2^t})) \text{ MOD } 2 = 1 \\ 0 & (\text{Int}(\frac{I(i, j)}{2^t})) \text{ MOD } 2 = 0 \end{cases} \quad (3-63)$$

式中， $0 \leq t \leq 7$ ， $B_t(\bullet)$  代表图像的位平面分解操作，显然  $g_t(i, j) \in \{0, 1\}$ 。经过这样的处理后，该图像就被分解为  $g_0, g_1, \dots, g_7$ ，从而就完成了对图像进行位平面分解的操作，如图 3.18 所示。反过来，图像像素  $I(i, j)$  的合成可以表示为

$$F(i, j) = \sum_{t=0}^7 g_t(i, j) \times 2^t \quad (3-64)$$

对于每个像素点  $I(i, j)$  来说， $g_7(i, j)$  反映了该点处灰度的最重要的情况，然后依次是  $g_6(i, j)$ ， $g_5(i, j)$ ， $\dots$ ， $g_0(i, j)$ 。

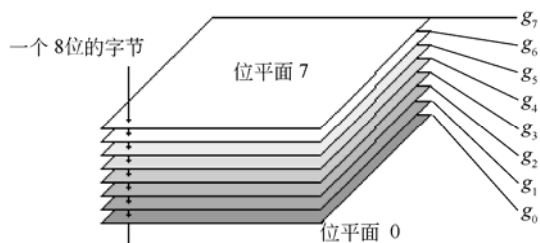


图 3.18 位平面分解示意图

一个 8 位灰度图像的 8 个位平面分解结果如图 3.19 所示，其中，图 3.19 (a) 是一幅示例图像，图 3.19 (b) 到 (i) 是它的 8 个位平面（从高位面到低位面）。从图中可以看出，每个位平面都能够反映图像的频率和方向在局部范围内的变化强度，较高位（特别是前 4 位）包含了大多数在视觉上很重要的数据，其他位平面则包含了图像中更多的细节信息。而且低位面图比高位面图复杂，包括的细节多，也更随机。把数字图像分解成位平面对于分析每一位平面在图像中的相对重要性是非常有用的，可以看出，图像的位平面存在一定的规律性，不同的位平面表现出不同的特性。由于相邻像素点颜色的相关性（即颜色值相差不大），较高的位平面有一定区域内值相同，随着权值的递减，这种区域越来越小，最后出现的是很多杂乱无章小区域和离散点。对于位平面而言，越高的位平面包含的信息就越重要，越低的位平面由于其杂乱无章而表现出一定的自相似性。而且，在高位平面出现随机纹理，在低位平面的相同位置仍出现随机纹理。即较不重要位平面的随机纹理多于较重要的位平面中的随机纹理。

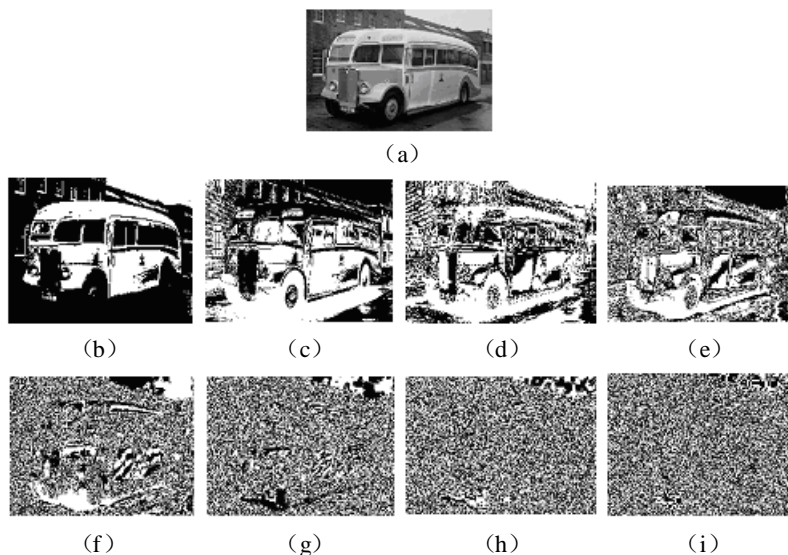


图 3.19 示例图像及其 8 位平面

采用这种位平面的表示方法存在一个缺点，即像素点灰度值的微小变化会对位平面的复杂度产生较明显的影响。例如，当空间相邻的 2 个像素的灰度值分别为 127 ( $01111111_2$ ) 和 128 ( $10000000_2$ ) 时，图像的每个位平面上在这个位置处都会有从 0 到 1（或从 1 到 0）的过渡。为减小这种影响，算法中采用灰度码的方法来表示位平面。 $m$  bit

图像的灰度码可由式 (3-65) 计算

$$G_i = \begin{cases} g_i \oplus g_{i+1}, & 0 \leq i \leq m-2 \\ g_i, & i = m-1 \end{cases} \quad (3-65)$$

式中,  $\oplus$  代表异或操作;  $g_i$  表示位平面分解得到的第  $i$  个位平面;  $G_i$  指位面  $g_i$  的灰度码表示。这种码的独特性质是相连的码字只有 1 个比特位的区别, 这样, 像素点灰度值的小变化就不会影响所有的位平面。而且对于每个用其相应的灰度码来表示的二进制位平面, 其灰度码是唯一的, 反之亦然。仍考虑上述空间相邻的 2 个像素的灰度值分别为 127 和 128 的例子, 如用式 (3-65) 的灰度码来表示, 这里只有位平面 7 有从 0 到 1 的一个过渡, 其他位平面没有变化。此时对应 127 和 128 的灰度码分别是  $01000000_2$  和  $11000000_2$ 。图 3.20 给出了图 3.19 中 (a) 图灰度码表示的位平面图 (从高位到低位)。可以看出, 用灰度码表示的位面图复杂度较低, 但具有视觉意义信息的位面图数量更多。

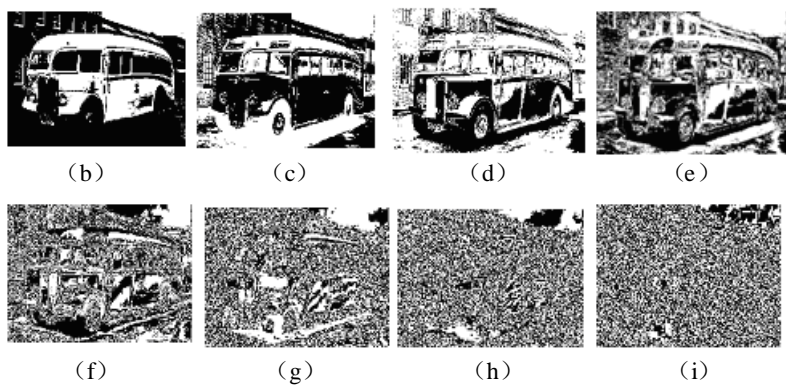


图 3.20 利用灰度码表示的位平面图

### 3.7.2 位平面熵

由前面的分析可知, 对于一幅图像通过位平面分解得到的几个位平面来说, 越高的位平面包含的对视觉理解重要的数据越多。由图 3.20 也可以看出, 仅仅使用一幅图像的高位部分图 3.20 (b) 就可以很好地表示整幅图像的内容, 只有一些要求比较高的特殊情况下才会用到图像中的低位部分。但是, 由于图像在不同的位平面有着不同的分布特性, 如果只选择其中的一个位平面, 不能对图像进行很好的描述。从图 3.19 和图 3.20 可以看出, 只有最高的几个位平面包含了视觉可见的有意义信息, 体现了明显的图像结构特征, 其他低位平面的信息随机性很强, 是很局部的小细节, 只增加图像的亮度信息, 没有提供任何的结构信息。因此, 这里仅采用图像高位的 4 个平面来提取图像特征, 具体过程如图 3.21

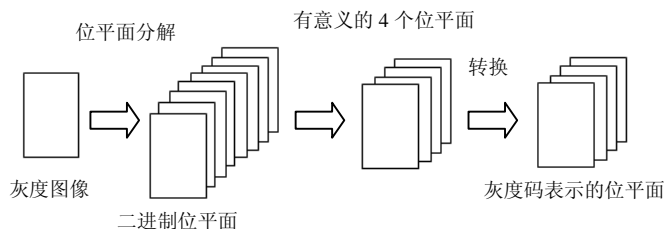


图 3.21 位平面分解及其灰度码表示



所示。由于位平面是一系列二值图像，对原始图像灰度值的变化非常敏感，噪声对其有较大的影响，因此，在进行特征提取之前，需要对图像进行消噪处理。

在此基础上，采用信息熵来描述每一位平面的特征。首先对图像通过位平面分解得到最高的 4 个位平面  $(g_7, g_6, g_5, g_4)$ ，然后转换成它们对应的灰度码表示的位平面  $(G_7, G_6, G_5, G_4)$ ，利用熵的定义计算每个位平面的信息熵。由于位平面为二值图像，灰度的取值只有 0 或 1 两种，所以位平面的信息熵 BE (Bit-plane Entropy) 可简化为

$$BE = -p_1 \log p_1 - p_0 \log p_0 \quad (3-66)$$

式中， $p_1$  和  $p_0$  代表位平面中值为 1 及 0 的像素出现的概率，且  $p_0 + p_1 = 1$ 。

由式 (3-66) 可以看出，如果仅仅对一幅二值图像来说，其信息熵只与图像中 0 和 1 像素的像素数有关，而与像素的空间位置无关，只利用单个位平面的信息熵作为特征，不能对图像进行很好的区分。然而这里利用四个对表征图像内容有意义的位平面，构造位平面熵矢量，因此，对具有相同信息熵而分布特征不同的两幅图像来说，它们的位平面分解得到的几个位平面却不一定相同，从而提取位平面熵矢量也不一样。如图 3.22 所示，图 (a) 和 (b) 两幅图像具有相似的颜色直方图 (Pass G. et al 1996)，但如果用采样位平面熵矢量表示，图 (a) 和 (b) 分别为  $\{0.668, 0.435, 0.517, 0.452\}$  和  $\{0.621, 0.385, 0.411, 0.416\}$ ，显然，通过二者的熵矢量可很好地进行区分。



图 3.22 示例图像

位平面熵和图像的全局信息熵一样，具有旋转不变性、尺度不变性和平移不变性，对图像的几何形变具有很强的鲁棒性。同时，由于各个位面实际上是提取了原图像中各个像素的灰度值的某一位形成的，因此，避免了直方图中颜色量化所带来的问题，而且也包含了图像像素的空间分布信息，解决了熵相同而图像不同的问题。熵矢量的计算只与每个位平面内图像像素点的数目有关，克服了原图像相邻像素点之间的相关性，忽略了图像的细微变化，即整体上求大同，局部上存小异。位平面熵的计算复杂度较低且维数较少，可以满足图像检索中存储容量和检索速度的要求。

### 3.8 位平面熵增强法

位平面是一种在图像编码中常用的方法，文献 (赵珊等 2007, 王向阳等 2007) 将图像位平面应用到了基于内容的图像检索中，采用图像较高位平面的特征，作为图像索引。由于图像较高位平面包含了图像主要内容，图像细节相当少、忙乱或随意性较小，而较低位平面所携带的原始信号能量极少，忙乱或随意性较大，因此忽略较低位平面对原图像质量没有太大影响。另外，图像噪声的随意性较大，也就是说噪声主要集中在较低位平面，

对图像较高位平面信息的影响很小,因此,提取较高位平面信息来刻画和表示图像内容不仅可以较好地保留原图像信息,而且还能够有效避免噪声干扰。

然而,3.7节在利用位平面及信息熵进行图像检索时,忽视了两个重要问题:①熵的对称性对检索效果的影响;②没有考虑位平面的空间信息。本节我们给出了相应的解决办法,以避免熵的对称性对利用位平面熵进行图像检索的性能造成的影响,同时在位平面特征中引入了空间信息,进一步增强了算法的检索性能(孙君顶等 2009)。

### 3.8.1 改进的位平面熵

这里仍采用灰度码表示的位平面来提取位平面熵(赵珊等 2007),这种编码可以保证连续码字只在一比特位置上不相同,因此,灰度级小的变化不太可能影响到所有的位平面。

按照 3.7.2 中的方法,位平面熵表示为  $-p_1 \log p_1 - p_0 \log p_0$  (式 (3-66)),  $p_1$  和  $p_0$  仍代表位平面中值为 1 及 0 的像素出现的概率。然而,根据熵的对称性可知,矢量各分量的次序任意改变时,熵值不变,熵函数的取值只与概率分布有关(见 3.3.8 节)。如图 3.23 所示,图 (a)、(b) 表示两幅视觉效果完全不同的图像,图 (c)、(d) 为图 (a)、(b) 的第 7 位平面示意图,图 (e)、(f) 为图 (c)、(d) 的统计直方图。可以看出,它们的直方图差别很大,但如果按照 3.7 节的方法,根据熵的对称性,这两个直方图具有相似的信息熵,这样,采样位平面熵就较难区分两个位平面。而且,随着图像库中图像数量的增多,出现这种情况的概率会越来越大,也就是说,随着图像库图像数量的增多,位平面熵的检索准确度和检索率都会下降。

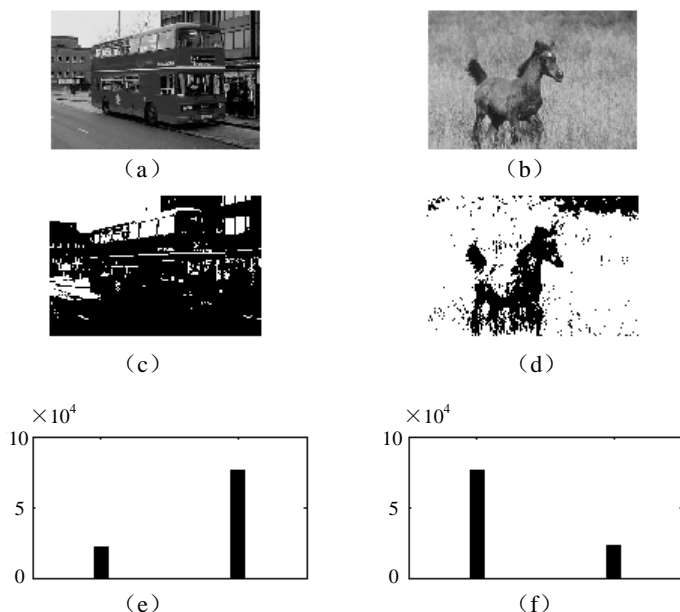


图 3.23 (a)、(b) 为两幅原始图像, (c)、(d) 为 (a) 及 (b) 的位平面 7, (e)、(f) 为 (c) 及 (d) 的直方图

为了解决该问题,这里给出了改进位平面熵(Enhanced Bit-plane Entropy, EBE),其定义如下

$$EBE = \begin{cases} BE, & \text{if } p_1 > p_0 \\ -BE, & \text{else} \end{cases} \quad (3-67)$$

这样,对于图 3.23 所示的位平面 7 来说,它们的熵就有正负之分,因此可以很容易进行区分。

但上述定义还存在一个问题。如图 3.24 所示,图 (a)、(b) 为两幅视觉效果相似的图像,图 (c)、(d) 为图 (a)、(b) 的第 7 位平面示意图,图 (e)、(f) 为图 (c)、(d) 的统计直方图。可以看出,它们具有相似的直方图,而且  $p_1$  和  $p_0$  的值也比较接近。但很明显,对于图 (e)  $p_1 > p_0$ ,对于图 (f)  $p_1 < p_0$ 。在这种情况下,如果还采用式 (3-67) 计算位平面熵,它们的熵仍是一正一负,这将导致检索错误。考虑到该问题,对 EBE 做了进一步的定义,即

$$EBE = \begin{cases} BE, & \text{if } p_1 > p_0 \text{ and } p_1/p_0 < \alpha \\ -BE, & \text{else} \end{cases} \quad (3-68)$$

$\alpha$  是事先设定的阈值。也就是说,在式 (3-67) 的基础上,如果  $p_1$  和  $p_0$  的比值小于给定的阈值,也就是说在  $p_1$  和  $p_0$  比较接近的情况下,直接按照 3.7 节的方法计算,否则按照新的定义计算。这就解决了当两幅图像直方图近似,而不同时存在  $p_1 > p_0$  或  $p_1 < p_0$  造成熵有正负之分的问题。

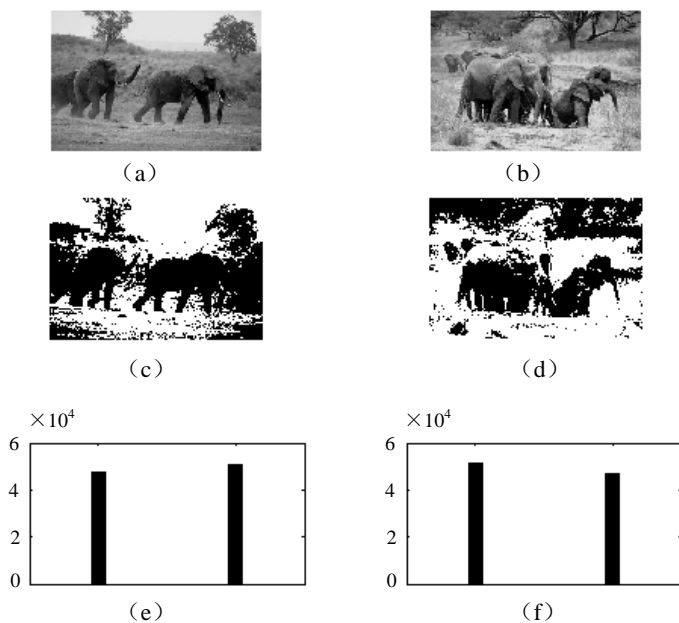


图 3.24 (a)、(b) 为两幅原始图像, (c)、(d) 为 (a) 及 (b) 的位平面 7, (e)、(f) 为 (c) 及 (d) 的直方图

### 3.8.2 空间分布熵

在 3.7 节,我们仅仅介绍采用位平面熵作为位平面的特征,并没有考虑位平面的空间特征,因此可能造成误检。如图 3.25 所示,图 (a)、(b) 为两幅视觉效果完全不同的图像,图 (c)、(d) 为图 (a)、(b) 的第 7 位平面示意图,图 (e)、(f) 为图 (c)、(d) 的

统计直方图。可以看出，两幅图像的位平面 7 具有十分相似的直方图，相近的位平面熵。但从图中可明显看出，虽然它们的直方图相似，但具有不同值的像素（黑、白两类）的分布差别却很大，也就是说它们具有不同的空间特征。因此，空间特征也是影响检索准确度的重要因素。这里，我们采用位平面分布熵来描述位平面的这种空间特征。

位平面分布熵的计算方法同空间分布熵的计算方法（可参阅 3.6 节）。设归一化后的环形直方图为  $(p_{i1}, p_{i2}, \dots, p_{iL})$ ，其中  $p_{ij} = |A_{ij}| / |A_i|$ 。结合信息熵概念，则位平面分布熵（Bit-plane Spatial Entropy, BSE）可表示为

$$BSE_i = -\sum_{j=1}^L p_{ij} \log_2(p_{ij}) \quad (3-69)$$

位平面分布熵反映了值为  $i$  像素的空间分布特征，且具有尺度、平移不变性及旋转不变性。由于位平面为二值图像，像素值为 0 或 1，这里我们取位平面分布熵 BSE 为

$$BSE = \max\{SE_0, SE_1\} \quad (3-70)$$

对于位平面分布熵，仍需要考虑熵的对称性的影响，仍可采用 3.4.9 节的改进方法来消除熵的对称特性对位平面分布熵的影响。

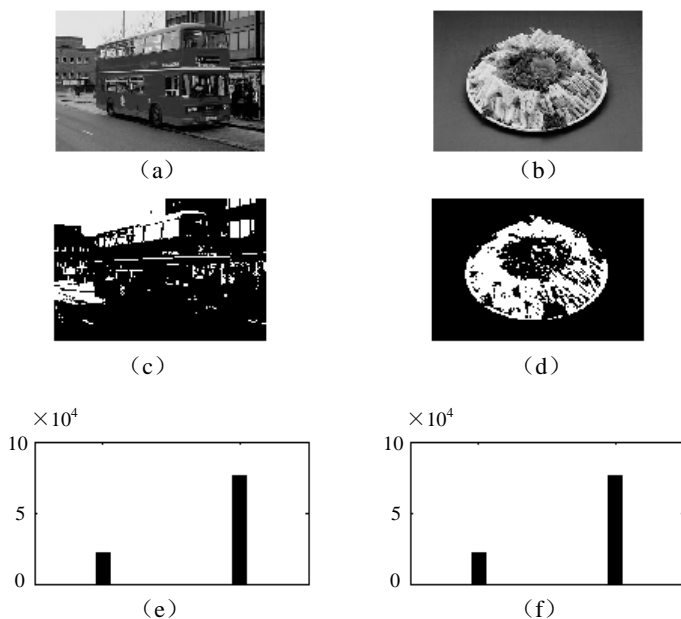


图 3.25 (a)、(b) 为两幅原始图像，(c)、(d) 为 (a) 及 (b) 的位平面 7，(e)、(f) 为 (c) 及 (d) 的直方图

### 3.9 基于显著点的图像检索

兴趣点作为一种重要的图像视觉特征，广泛用于三维解释、运动估计以及图像匹配等方面（Beaudet P.R. 1978, Kitchen L. et al 1982, Harris C. et al 1998）。传统的兴趣点检测方法，往往使用基于数学模型的角检测器，这类兴趣点又称为角点。近几年来，兴趣点被广泛用于图像检索（Bres S. et al 1999, Han J.W. et al 2002）。然而，传统的角检测器存在两个主要缺陷：①兴趣点并不能完全代表图像检索中最感兴趣的索引点，索引点应该在任

何视觉上最感兴趣的地方，而不管它是位于角点的位置还是光滑的边缘上。因此，视觉焦点不一定是角点，光滑的边也可能是视觉焦点，但传统的角检测器却检测不到这些视觉焦点。②采用传统角检测器检测到的角点可能仅仅聚集在图像中的某些小区域，如果仅根据角点来提取图像的特征，那么图像中除角点之外的其他区域的特征就无法描述。

为了解决利用兴趣点进行检索时存在的问题，显著点的概念 (Sebe N. et al 2000) 被提了出来。显著点主要是指图像中一些相对稳定而突出的点，因此，它不仅包括传统的图像兴趣点，且将所有能引起视觉注意的感兴趣区域的一些信息点也包括在内。可以看出，显著点是更广泛意义下的兴趣点。目前，针对显著点的提取算法主要是利用小波变换在原图像上直接提取，这些方法得到的显著点在图像中的分布比较分散，并且所采用的提取算法也比较复杂 (Sebe N. et al 2000, Tian Q. et al 2001)；同时，这些算法在特征提取时仅考虑了显著点的灰度变化特征，忽视了显著点作为整体所暗含的形状结构特征。本节在图像的块逆概率差 (Block Difference of Inverse Probabilities, BDIP) (Yong D.C. et al 2003) 模型的基础上，得到原图像的 BDIP 图像，然后根据 BDIP 图像中像素的分布特点来提取图像的显著点，然后以它们为线索，把图像的形状特征和空间颜色分布特征有机的结合起来进行检索。

### 3.9.1 块逆概率差 (BDIP) 模型及BDIP图像的提取

逆概率差 (difference of inverse probabilities, DIP) (Ryoo Y.J. et al 1988) 是一种提取与图像灰度值相关的、包含脊和边缘等骨架特征的模型。在这种模型中，一个图像窗口中某一个像素的灰度值与窗口中全体像素的灰度值总和的比被看做一个概率事件。因此，从这个意义上来说，DIP 就意味着图像窗口中中心像素的逆概率与窗口中最大灰度值的像素的逆概率之差。在此基础上，文献 (Yong D.C. et al 2003) 又提出使用图像的块逆概率差 (BDIP) 作为图像的纹理描述符并用于图像检索。具体定义如下：

$$BDIP = M^2 - \frac{\sum_{(i,j) \in B} I(i,j)}{\max_{(i,j) \in B} I(i,j)} \quad (3-71)$$

式中， $I(i, j)$  表示像素点  $(i, j)$  处的灰度值， $B$  表示大小为  $M \times M$  的图像块。

用图像块的 BDIP 值表示的图像，称为原图像的 BDIP 图像。在利用这个模型提取图像的显著点时，由于 BDIP 图像中像素值的分布比较分散，因此需要对其进行归一化处理，归一化的 BDIP 值定义如下：

$$BDIP = \frac{\left( M^2 - \frac{\sum_{(i,j) \in B} I(i,j)}{\max_{(i,j) \in B} I(i,j)} \right)}{M^2} \quad (3-72)$$

对一幅真彩色图像来说，在提取 BDIP 图像时要考虑选择一个合适的颜色空间，提取的 BDIP 图像才能更好地刻画原始图像的内容特征。由第 2 章可知，在图像检索中常用的颜色空间包括 RGB、HSV、CIEL\*a\*b\* 和 CIEL\*u\*v\* 等。一般认为，RGB 颜色空间与人眼的感知差别很大，而 HSV 颜色空间是一种面向视觉感知的颜色模型，直接对应于人眼色彩视觉特征的三要素，能很好地与人眼的视觉特性相匹配。同时，由于 HSV 空间各轴在视觉上彼此无关，并且从 RGB 到 HSV 的转换比较简单，所以，这里采用较好符合人眼感知特性的 HSV 颜色空间作为对彩色图像进行处理的颜色空间。HSV 颜色空间有两个主要

特点：其一，亮度分量与图像的彩色信息无关；其二，色调和饱和度分量与人感受颜色的方式是紧密相连的。色调  $H$  和饱和度  $S$  反映了颜色的种类和颜色的深浅程度，在人类对颜色的视觉判断中起着主要的影响作用。在基于内容的图像检索中，应用这种颜色模型会更适合用户的视觉判断。

理论上讲，在提取图像的颜色特征时使用真彩色将产生最小的误差，但是此时特征提取和特征匹配所需的计算量以及特征的存储所需的存储空间都将随实际颜色数目增多而非线性膨胀。实际上，对图像颜色进行一定的量化，尽管图像质量会有所下降，但并不影响人们对图像内容的正确理解。在对图像进行处理时，先将图像从  $RGB$  空间转换到  $HSV$  空间，对于图像颜色的量化，这里采用黄元元（黄元元 2003）提出的量化方法，将图像的颜色量化为 36 柄。量化过程如下：

$$H = \begin{cases} 0, & H \in [0^\circ, 60^\circ) \\ 1, & H \in [60^\circ, 120^\circ) \\ 2, & H \in [120^\circ, 180^\circ) \\ 3, & H \in [180^\circ, 240^\circ) \\ 4, & H \in [240^\circ, 300^\circ) \\ 5, & H \in [300^\circ, 360^\circ) \end{cases} \quad S = \begin{cases} 0, & S \in [0, 0.25) \\ 1, & S \in [0.25, 1] \end{cases} \quad V = \begin{cases} 0, & V \in [0, 0.3) \\ 1, & V \in [0.3, 0.8) \\ 2, & V \in [0.8, 1.0] \end{cases} \quad (3-73)$$

根据式 (3-73) 的划分，可以将很多虽然深浅不同但在视觉上仍属于同一类的颜色量化在同一区间内，使量化结果更加符合人类的视觉感受。量化完成后，将  $HSV$  颜色空间划分为  $n_H \times n_S \times n_V$  个区间，其中  $n_H$ 、 $n_S$  和  $n_V$  分别表示  $H$ 、 $S$  和  $V$  三个分量的量化级数，即  $n_H = 6$ ， $n_S = 2$ ， $n_V = 3$ 。也就是说，将  $HSV$  空间划分为 36 个区间。按照以上的量化级数，把  $H$ 、 $S$ 、 $V$  三个分量合并为一个一维特征向量，即

$$n = n_S \times n_V \times H + n_V \times S + I \quad (3-74)$$

式 (3-74) 又可写为

$$n = 6H + 2S + I \quad (3-75)$$

图 3.26 给出了用  $BDIP$  值表示的原图像，即图像的  $BDIP$  图像，子图像块的大小取为  $3 \times 3$ 。其中图 (a) 是原始图像，图像大小为  $384 \times 256$ ，图 (b) 是原图像的  $BDIP$  图像，图像大小为  $128 \times 85$ 。由图中可以看出，对于灰度变化明显的块，其  $BDIP$  值也会很大。而且  $BDIP$  图像中的像素点几乎都位于图像中能引起注意的视觉焦点位置，因此利用  $BDIP$  图像可以很好地表示图像的基本特征。

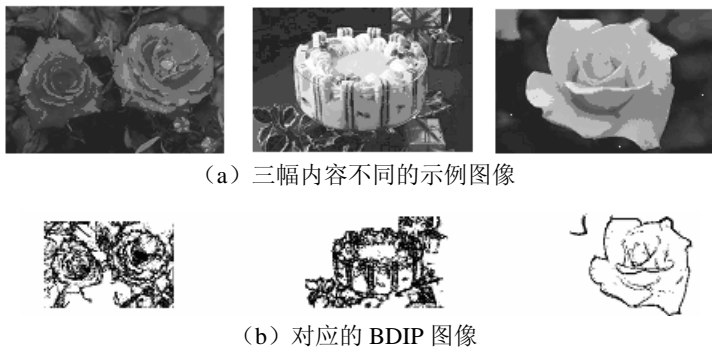


图 3.26 示例图像及其对应的  $BDIP$  图像

### 3.9.2 显著点提取算法

目前,针对显著点的提取算法主要是利用小波变换在原图像上直接提取,这些方法得到的显著点在图像中的分布比较分散,并且所采用的提取算法也比较复杂。这里,利用BDIP图像进行显著点提取,选择它作为处理对象使得计算量大为减少。此外,BDIP图像的对比特较原图像有所提高,检测显著点也较为容易。在提取图像的BDIP图像时发现,图像的BDIP值与相对应的图像块的信息变化有密切的关系,BDIP值越大,表示对应图像块中信息越丰富,即纹理边缘信息越多,表征图像的能力越强。利用这个特点进行图像显著点的提取。

对一幅原图像变换后得到的BDIP图像来说,其中的像素点能完整的表示出图像的内容特征,考虑到这点,在算法中将BDIP图像中灰度值不为0的像素点先作为显著点的初选。这些显著点并不都是重要的,我们要选取一些突出的足够引起视觉注意的显著点。应该考虑到,所选取的显著点个数对图像的检索效果有很大的影响,显著点选取的较少,体现不出图像的特征,若选取的过多,会增加图像特征提取的计算量。文献(Tian Q. et al 2001)在处理这个问题时,仅仅简单地取前 $N$ 个较大的显著点并以其为基础提取图像的相关特征。但 $N$ 的值并不适合所有的图像。同样,若给定一个门限值,也不能适合所有的图像。通过实验,发现图像块中的灰度变化程度不仅可以用图像的BDIP值来反映,还可用块内灰度值的方差来体现。方差的大小和BDIP值的大小是一致的。BDIP值大,块内的方差也大,反之亦然。因此结合块内灰度值的方差 $\sigma$ ,给出一个有一定自适应性的显著点选择条件,即

$$\sigma \times V_{BDIP} \geq \lambda \mu_T \quad (3-76)$$

式中, $V_{BDIP}$ 为图像块的BDIP值, $\lambda$ 为实验中需设定的参数, $\mu_T$ 是整幅图像中图像块BDIP值的均值。这样,一幅图像中显著点的选择就可以根据自身图像块的方差、BDIP值及整幅图像中图像块BDIP值的均值决定。

图3.27给出了几幅图像的BDIP图像及提取的显著点,各个图像的分辨率和大小与图3.26相同。从图中可以看出,式(3-76)中的条件倾向于选取块内方差较大即BDIP值较大的显著点,而且采用该算法不仅可以检测出角点,而且还能检测出平滑边缘上的点。图中显著点都集中在图像中视觉焦点上,而且显著点构成的轮廓可以清晰地描述图中物体的形状。边缘点和角点都是引起人视觉注意的地方,在这些地方提取图像特征,有利于图像检索结果和人的视觉保持一致。

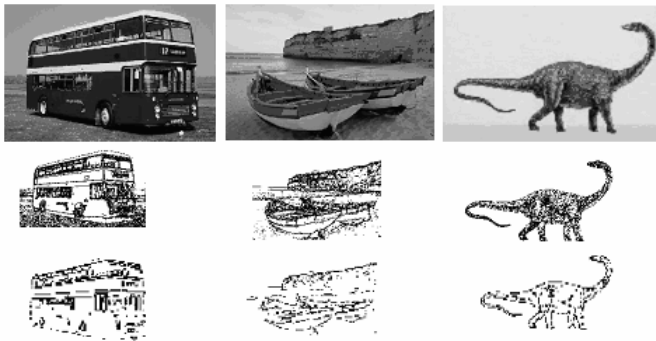


图 3.27 示例图像及相应的BDIP图像和显著点提取图像

图 3.28 给出了一幅图像在不同状态下的显著点提取图。其中，图 (a) 是示例图像，图 (b) 是加噪图像，图 (c) 是旋转一定角度的图像，图 (d) 是平移后的图像。由图中可以看出，由于本方法是以图像块为基本单位的，噪声对一个像素点的影响可能很大，但对一个块来说影响却不是很大，因此采用本文算法提取的显著点对噪声有很强的抵抗力。同时，也可以看出，经过旋转和平移后提取的显著点基本上变化不大。所以该算法具有一定的鲁棒性。

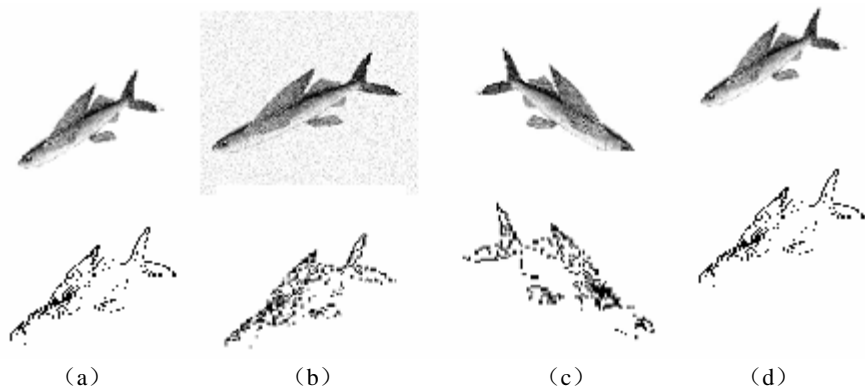


图 3.28 显著点鲁棒性实验

图 3.29 给出了 BDIP 方法与 Han 方法 (Han J.W. et al 2002) 及 Tian 方法 (Tian Q. et al 2001)，其中，图 (a) 是示例图像，图 (b) 是 Han 方法提取的兴趣点，图 (c) 是 Tian 方法提取的显著点，图 (d) 是 BDIP 方法提取的显著点。

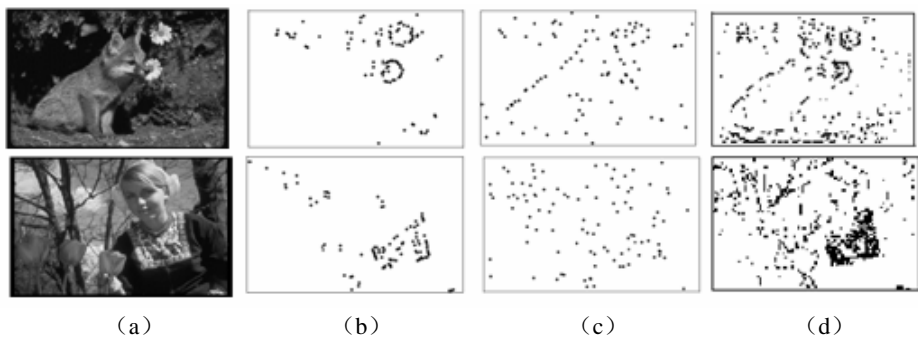


图 3.29 不同方法提取特征点示意图

### 3.9.3 基于显著点的特征提取

显著特征点与图像中视觉上最重要的部分相关联，根据它提取的图像特征具有很强的表征能力；同时，由于显著点定位于一些可视焦点上，能抓住图像的局部信息，用显著特征点周围的图像信息代替全局图像信息能减少算法的计算复杂度。另外，从显著点的提取过程来看，它们对于噪声等的影响具有较强的鲁棒性。因此，充分利用图像的显著点，有利于提高图像的检索效率。

由于显著点往往分布在图像中视觉上最感兴趣的地方，蕴含了丰富的颜色细节，所以



利用显著点局部区域的颜色特征来描述图像具有合理性。每个显著点所对应的 BDIP 值在一定程度上可以体现其与周围点的灰度对比度, BDIP 值越大, 显著点对应的灰度值也越大。因此可以结合 BDIP 值的分布特征提取显著点的颜色空间分布特征。

为了提取颜色的空间分布信息, 赵珊(赵珊 2007)对显著点的 BDIP 值的分布特征进行了统计, 所选取的图像被划分为  $3 \times 3$  大小的子块, 实验结果表明, 大部分显著点的值在  $[0, 0.1]$ , 少量的值分布在  $[0.1, 0.4]$ , 其余的只有极少一部分取值在  $[0.4, 1]$ 。

根据上述显著点的分布, 可以将其分成三种状态: 密集态、常态和稀疏态。通过提取这三种状态下显著点颜色分布的一阶矩和二阶矩来描述颜色及其空间分布信息。定义如下:

$$\begin{aligned}\mu_p &= \frac{1}{N_p} \sum_{V_{BDIP} \in p} V_{BDIP} \\ \sigma_p &= \sqrt{\frac{1}{N_p} \sum_{V_{BDIP} \in p} (V_{BDIP} - \mu_p)^2}\end{aligned}\quad (3-77)$$

式中,  $p=0, 1, 2$ , 分别表示显著点的三种状态,  $V_{BDIP}$  为显著点的 BDIP 值,  $N_p$  表示属于  $p$  状态下显著点的数目,  $\mu_p$ 、 $\sigma_p$  分别表示  $p$  状态下显著点 BDIP 值分布的一阶矩和二阶矩。

## 参 考 文 献

- [1] Androustos D., Plataniotis K.N., Venetsanopoulos A.N., 1997. Image retrieval using the directional detail histogram. SPIE 3312, 129-137.
- [2] Arthur R., Weeks G., 1997. Color segmentation in the HIS color space using the k-means algorithm. SPIE, 9(6): 143-154.
- [3] Beaudet P.R., 1978. Rotationally invariant image operators. International Conference on Pattern Recognition, 579-583.
- [4] Bres S., Schettine R., 1999. Detection of interest points for image indexation. IEEE conference on image processing, 227-234.
- [5] 陈宗海, 文锋, 聂建斌等. 基于结点生长k-均值聚类算法的强化学习方法. 计算机研究与发展, 2006, 43(4): 661-666.
- [6] Cinque L., Levialdi S., Olsen K., et al, 1999. Color-based image retrieval using spatial-chromatic histograms. In Multimedia Computing and System, IEEE Intl Conf. 2: 969-973.
- [7] Constantium N., Nozha B., 2000. Spatially constrained color distributions for image indexing. International Conference on Color in graphics and image processing (CGIP).
- [8] Dimai A., 1997. Spatial encoding using differences of global features. In Proc. of SPIE storage and retrieval for image and video database IV, 3022: 352-360.
- [9] Dony R.D., Wesolkowski S., 1999. Edge detection on color images using RGB vector angles. In: Proceedings of IEEE Canadian Conference on Electrical & Computer Engineering, 687-692.
- [10] Finlayson G.D., Chatterjee S.S., Funt B.V., 1996. Color Angular Indexing, ECCV96, II: 16-27.
- [11] Finlayson, G., Hordley, S., Schaefer, G., et al, 2005. Illuminant and device invariant color using histogram equalisation. Pattern Recognition, 38: 179-190.
- [12] Funt B.V., Finlayson G.D., 1995. Color constant color indexing. IEEE Transactions on Pattern Analysis and Maching Intelligence, 17: 522-529.
- [13] 高新波. 模糊聚类分析及其应用. 西安: 西安电子科技大学出版社, 2004.

- [14] 高新波, 李洁, 姬红兵. 基于加权模糊 c 均值聚类与统计检验指导的多阈值图像自动分割算法. 电子学报, 2004, 32(4): 661-664.
- [15] 耿国华. 周明全. 常用色彩量化算法的性能分析. 小型微型计算机系统, 1998, 19(9): 46-49.
- [16] Gerrautz M., Purgathofer W., 1998. A simple method for color quantization: octree quantization. Proc of ICG'98, 8(6): 219-230.
- [17] Gevers T., Smeuder A.W.M., 1996. Evaluating color and shape invariant image indexing of consumer photograph. In: Proceedings of the 1<sup>st</sup> International Conference on Visual Information Systems, Melbourne, Australia, 254-261.
- [18] Gevers T., Smeuder A.W.M., 1999. Content-based image retrieval by viewpoint-invariant image indexing. Image and Vision Computing, 17(7): 475-488.
- [19] Gevers T., Smeulders A.W.M., 2000. PicToSeek: combining color and shape invariant features for image retrieval. IEEE Transactions on Image Processing, 9: 102-119.
- [20] Gevers T., 2001. Robust histogram construction from color invariants. ICCV01, I: 615-620.
- [21] Gevers T., Stokman H.M.G., 2004. Robust histogram construction from color invariants for object recognition. PAMI, 26(1): 113-118.
- [22] Han J.W., Guo L., 2002. New image retrieval approach based on interest points. SPIE, 4862: 197-197.
- [23] Harris C., Stephens M., 1998. A combined corner and edge detection. Image vision computing, 6: 121-128.
- [24] 何海涛, 许端清, 谭彬. 一种基于 K-means 的色彩量化快速算法. 中国图像图形学报, 2003, 8(A): 767-770.
- [25] 何清法, 李国杰. 综合分块主色和相关反馈技术的图像检索方法. 计算机辅助设计与图形学学报, 2001, 13(10), 912-917.
- [26] Healey G., Slater D., 1994. Global color constancy: recognition of object by use of illumination-invariant properties of color distributions. Journal of the Optical Society of America A, 11(11): 3003-3010.
- [27] Hsu W., Chua T. S., Pung H. K., 1995. An integrated color-spatial approach to content-based image retrieval. In: Proc ACM Multimedia'95 Conference, San Francisco: 305-313.
- [28] Huang J., 1998. Color-spatial image indexing and applications. Phd. thesis, Cornell University.
- [29] 黄元元. 基于视觉特征的图像检索技术研究. 博士学位论文, 南京理工大学, 2003.
- [30] Ji Y.K., Chang Y.K., Yang S.S., et al, 2001. Color indexing using chromatic invariant. Pattern Recognition 34: 1189-1197.
- [31] 金振玉. 信息论. 北京: 北京理工大学出版社, 1991.
- [32] John Z M, 2000. An Information theoretic approach to content based image retrieval. Phd. Thesis, Louisiana State University and Agricultural and Mechanical College.
- [33] Kanungo T., Mount D.M., Netanyahu N., et al, 2002. Efficient k-means clustering algorithm: analysis and implementation, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7): 881-892.
- [34] Kenneth.R.Castleman. 数字图像处理. 北京: 电子工业出版社, 1998.
- [35] Kitchen L., Rosenfeld A., 1982. Gray-level corner detection. Pattern recognition letters, 1: 95-102.
- [36] Krishnan N., Sheerin M.B., Christiyana C., 2007. Content Based Image Retrieval Using Dominant Color Identification Based on Foreground Objects. Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), Volume 03: 190-194.
- [37] Li J., Wang J.Z., Wiederhold G., 2000. IRM: integrated region matching for image retrieval. Proceeding of ACM Multimedia, 2000.
- [38] 李向阳, 庄越挺, 潘云鹤. 基于内容的图像检索技术与系统. 计算机研究与发展, 2001, 38(3): 344-354.
- [39] 梁艳梅, 翟宏坤, 毋国光. 基于模糊相关的彩色图像检索. 中国科学 (E 辑), 2003, 33(10): 934-938.
- [40] Lim S., Lu G. J., 2003. Spatial statistics for content-based image retrieval. Proceedings of the International Conference on Information Technology: Computers and Communications(ITCC).

- [41] Lim Y.W., Lee S.U., 1990. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recognition*, 15(9): 935-952.
- [42] 林福宗. 多媒体技术基础 (第二版). 北京: 清华大学出版社, 2002.
- [43] 刘忠伟. 利用局部累加直方图进行彩色图像检索. *中国图像图形学报*, 1998, 3(7): 533-537.
- [44] 卢汉清. 图像视频信号的浏览与检索. *中国图像图形学报*, 2000, 5(B)(1): 5-9.
- [45] Lu G., Teng S., 1999. A novel image retrieval technique based on vector quantization. *Computational Intelligence for Modeling Control and Automation*, Viana, Austria, 160-165.
- [46] Manjunath B.S., Ohm J. R., Vasudenvan V. V., et al, 2001. Color and texture descriptors. *IEEE Trans. On CSVT.*, 11(6): 703-715.
- [47] 孟庆生. 信息论. 西安: 西安交通大学出版社, 1986.
- [48] Mathias E., Conci A., 1998. Comparing the influence of color spaces and metrics in content-based image retrieval. in: *International Symposium on Computer Graphics, Image Processing, and Vision*, 371-378.
- [49] Mojsilovic A., Hu J., Soljanin E., 2002. Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis, *Trans. Image Process.* 11 (11): 1238-1248.
- [50] Niblack W., Barber R., Equitz W., et al. 1993. The qbic project: Quering images by content using color, texture, and shape. In *Pocceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, 2-3 February '93, San Jose, CA, 173-187.
- [51] Novak C.L., Shafer S.A. 1991. Supervised color constancy for machine vision. In: *Human Vision. Visual Processing and Digital Display II*, SPIE 1453: 353-368.
- [52] Oleg V., 1995. Color Image quantization in windows systems with local K-means algorithm. *Proceeding of VI Western Computer Graphics Symposium*. 74-79.
- [53] Pass G, Zabin R, Miller J., 1996. Comparing Images Using Color Coherence Vectors. In *ACM International Conference on Multimedia*. Boston, MA, 65-73.
- [54] Qiu G.P., Lam K.M., 2003. Frequency layered color indexing for content-based image retrieval. *IEEE Transactions on Image Processing*, 12(1): 102-113.
- [55] Ravishankar K.C. , Prasad B.G , Gupta S. K. , et al, 1999. Dominant Color Region Based Indexing for CBIR. *ICIAP*, 887-902.
- [56] Rao A., Srihari R., Zhang Z., 1999. Spatial Color Histogram for Content-Based Retrieval. In *11<sup>th</sup> IEEE International Conference on Tools with AI*, 183-186.
- [57] Ryoo Y.J., Dim N.C., 1988. Valley operator extracting sketch features: DIP. *Electron. Lett.*, 4(248): 461-463.
- [58] Schettini R., Ciocca G., Zuffi S., 2001. A survey of methods for color image indexing and Retrieval in Image Databases. *Color Imaging Science: Exploiting Digital Media*, 183 -211.
- [59] Scheunders P. 1996. A genetic approach towards optimal color image quantization. *Image Processing*, 7(5): 1031-1034.
- [60] Sciascio E. D., Mingolla G., Mongiello M., 1999. Content-based image retrieval over the web using query by sketch and relevance feedback. In: *Proc of VISUAL'99*, 123-130.
- [61] Sebe N., Lew M. S., Huijsmans D.P., 1999. Multi-scale sub-image search. In *Proc. of 7<sup>th</sup> ACM Int. Conf. On Multimedia (Part 2)*, 79-82.
- [62] Sebe N., Tian Q., Loupias E et al, 2000. Color indexing using wavelet-based salient points. *IEEE workshop on content-based access of image and video libraries*, *IEEE CVPR'2000*, Hilton Head Island, South Carolina, June, 13-16.
- [63] Shannon, C. E., 1948. A mathematical theory of communication. *Bell Systems Technical Journal*, Vol. 27.
- [64] Shyh W.T., Lu G.J., 2007. Image indexing and retrieval based on vector quantization, *Pattern Recognition* 40: 3299-3316.
- [65] Smith J R, 1997. Integrated spatial and feature image systems: retrieval, analysis and compression. Ph.D Thesis, Columbia University.

- [66] Smith J.R., Chang S.F., 1997. Querying by color regions using the VisualSEEK content-based visual query system. In M. T. Maybury, editor, *Intelligent Multimedia Information Retrieval*. AAAI Press.
- [67] 宋擒豹, 杨向荣, 沈钧毅等. 图像相似模式挖掘中的颜色一位置直方图方法. *计算机研究与发展*, 2002, 39(9): 1132-1137.
- [68] Stehling R. O., Nascimento M. A., Falcao A. X., 2000. On 'shapes' of colors for content-based image retrieval. In the *ACM Multimedia Conference*. Los Angeles, USA, 171-174.
- [69] Stricker M., Orengo M., 1995. Similarity of color images. In: *Proceedings of SPIE Storage and Retrieval for Image and Video Database*, 2420: 381-392.
- [70] Stricker M., Dimai A., 1997. Spectral covariance and fuzzy regions for image indexing. *Machine Vision and Applications*, Vol 10, 66-73.
- [71] 孙君顶, 崔江涛, 毋小省等. 基于颜色和形状特征的彩色图像检索方法. *中国图像图形学报*, 2004a, 9(7): 820-827.
- [72] 孙君顶, 武学东, 周利华. 基于颜色和形状的图像检索. *计算机科学*, 2004b, 31(5): 180-183.
- [73] 孙君顶, 毋小省, 周利华. 基于信息熵的图像检索, *西安电子科技大学学报*, 2004c, 31(2): 223-228.
- [74] 孙君顶, 张喜民, 崔江涛等. 一种新的基于颜色和空间特征的图像检索方法. *计算机科学*, 2005a, 32(6): 158-160.
- [75] 孙君顶, 周利华. 一种改进的基于熵的图像检索算法. *红外技术*, 2005b, 27(1): 45-48.
- [76] 孙君顶, 丁振国, 周利华. 基于颜色及信息熵的彩色图像检索方法. *红外与毫米波学报*, 2005c, 24(2): 135-139.
- [77] 孙君顶. 基于内容的图像检索技术研究. 博士学位论文, 西安电子科技大学, 2005.
- [78] Sun J.D., Zhang X.M., Cui J.T., et al, 2006b. Image retrieval based on color distribution entropy, *Pattern Recognition Letters*, 27(10): 1122-1126.
- [79] Sun J.D., 2006. Image Retrieval Based on Improved Entropy and Moments, *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, USA.
- [80] 孙君顶, 崔江涛, 刘卫光等. 基于熵的图像空间特征提取及检索方法. *系统工程与电子技术*, 2006a, 28(6): 791-794.
- [81] 孙君顶, 毋小省. 基于分块主色和形状特征的彩色图像检索. *光电工程*, 2006, 33(12): 85-90.
- [82] 孙兴华. 基于内容的图像检索研究. 博士学位论文, 南京理工大学, 2001.
- [83] Swain M.J., Ballard D.H., 1991. Color indexing. *Intl. J. on Computer Vision*, 7(1): 11-32.
- [84] 汤海纓, 庄天戈. 计算机彩色模型在图像显示与分割中的应用. *计算机学报*, 1999, 22(41): 375-382.
- [85] Teng S., 2003. Image Indexing and retrieval based on vector quantization, Ph.D. Thesis, Monash University.
- [86] Teng S., Lu G., 2006. Efficient implementation of vector quantization for image retrieval, in: *The 12th International Multi-Media Modelling Conference*, Beijing, China, IEEE Press, New York, 280-287.
- [87] Tian Q., sebe N., Lew M.S. et al, 2001. Image retrieval using wavelet-based salient points. *Journal of Electronic Imaging*.
- [88] Van den Broek E.L., Schouten Th.E., Kisters P.M.F., 2008. Modeling human color categorization, *Pattern Recognition Letters* 29, 1136-1144.
- [89] Wang S.R., Chia L.T., Deepu R., 2003. Image retrieval using dominant colour descriptors. *International Conference on Imaging Science, Systems and Technology*.
- [90] 王涛, 胡事民, 孙家广. 基于颜色-空间的图像检索. *软件学报*, 2002, 13(10): 2031-2036.
- [91] 王炜, Michael R L, 武德峰等. 基于模糊分类的图像颜色直方图研究. *模糊系统与数学*, 2003, 17(4): 94-98.
- [92] 王向阳, 胡峰丽. 基于位平面颜色直方图的图像检索新方法. *小型微型计算机系统*, 2007, 28(4): 715-719.
- [93] 王宇生, 陈纯. 一种新的基于色彩的图像检索算法. *计算机研究与发展*, 2002, 39(1): 105-109.
- [94] 魏宝刚, 李向阳, 鲁东明, 潘云鹤. 彩色图像分割研究进展. *计算机科学*, 1999, 26(4): 59-62.

- [95] Xiuqi LI, Shu-Ching CHEN, Mei-Ling SHU, et al. 2002. A novel hierarchical approach to image retrieval using color and spatial information. Proceeding of The Third IEEE Pacific-Rim Conference on Multimedia (PCM2002), 16-18.
- [96] 徐旭, 朱淼良, 梁倩卉, et al. 一种用于 CBIR 系统的主色提取及表示方法. 计算机辅助设计与图形学学报, 1999, 11(5): 385-388.
- [97] 杨静宇, 曹雨龙. 计算机图像处理及常用算法手册. 南京: 南京大学出版社, 1997.
- [98] Yang N.-C., Chang W.-H., Kuo C.-M., et al, 2008. A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval. J. Vis. Commun. Image R.19 : 92-105.
- [99] 姚敏等编著. 数字图像处理. 北京: 机械工业出版社, 2006.
- [100] 姚琼, 赖剑煌, 冯国灿. 基于颜色特征检索方法的研究进展. 中国图像图形学报, 2003, 8(A): 418-423.
- [101] Yong D.C., Sang Y.S., Nam C.K., 2003. Image retrieval using BDIP and BVLC moments. IEEE transactions and systems for video technology, 13(9): 951-957.
- [102] 袁昕, 朱淼良. 基于主色匹配的图像检索系统. 计算机辅助设计与图形学学报, 2000, 12 (12): 917-921.
- [103] 袁昕, 吴春明, 朱淼良等. 一种用于 CBIR 系统的主色提取及表示方法. 计算机研究与发展, 2002a, 39(9): 1120-1126.
- [104] 袁昕, 朱淼良, 王东辉. 基于图像主色特征的 CBIR. 计算机辅助设计与图形学学报, 2002b, 14(11): 1061-1065.
- [105] Yoo H.W., Jang D.S., NA Y.K., 2002. An efficient indexing structure and image representation for content-based image retrieval. IEICE Trans INF&SYST, 1390-1398.
- [106] Yook K.L., 2004. Human perception based color image quantization. Proc of ICPR'04, 8(1): 664-667.
- [107] Zhang Y. J., Liu Z.W., He Y., 1997. Comparision and improvement of color-based image retrieval techniques. SPIE 3312, 371-382.
- [108] 章毓晋. 图像工程上册-图像处理与分析. 北京: 清华大学出版社, 1999.
- [109] 赵珊, 安志勇, 周利华. 基于颜色矢量角的彩色图像检索算法研究. 红外技术, 2006, 28(8): 460-465.
- [110] 赵珊. 基于内容的图像检索关键技术研究. 博士学位论文, 西安电子科技大学, 2007.
- [111] 赵珊, 崔江涛, 周利华. 一种基于图像显著特征点的检索算法. 模式识别与人工智能, 2007, 20(3): 382-387.
- [112] 赵珊, 崔江涛, 周利华. 一种新的基于颜色矢量角的图像空间描述符. 光电工程, 2007, 34(3): 77-81.
- [113] 赵珊, 崔江涛, 周利华. 基于位平面分布熵的图像检索算法. 电子与信息学报, 2007, 29(4): 795-799.
- [114] Zhou B., Shen J.Y., Peng Q.B., 2004. An adjustable algorithm for color quantization. Pattern Recognition letters, 25(16): 1787-1797.
- [115] Zhou B., Shen J.Y., Peng Q.K., 2004. An adjustable algorithm for color quantization, Pattern Recognition Letters 25, 1787-1797.
- [116] 孙君顶, 毋小省. 基于位平面熵及分布熵的图像检索. 系统工程与电子技术, 2009, 31 (3): 719-722.

## 第4章 基于形状特征的图像检索

形状特征是图像最主要的视觉特征之一，人们对一幅图像的理解很大程度上依赖于对图像中目标形状的区别和感知。因此，基于形状特征的图像检索也是 CBIR 中的重要研究内容。本章在介绍常用形状特征描述方法的基础上，详细讨论了基于状态矩阵的描述方法、基于平坦度和凹凸度的描述方法、基于信息熵的描述方法、基于链码的描述方法及基于矩的轮廓描述方法。

### 4.1 形状表达和描述

在计算机视觉中，相对颜色或纹理等低层特征而言，形状特征属于图像的中间层特征，它作为刻画图像中物体和区域特点的重要特征，是描述高层视觉特征（如目标、对象）的重要手段，而目标、对象对获取图像语义尤为重要。要把图像低层特征与高层特征有机地结合起来，必须有好的形状特征描述与提取算法的支持。由于物体形状的自动获取比较困难，基于形状的检索一般仅限于容易识别的物体（Faloutsos C. et al 1994, Pentland A. et al 1994）。对利用形状特征的图像检索，人们已提出了许多不同的方法，有关基于形状特征的图像检索算法可参阅相关综述性参考文献（丁险峰等 2001, Fudos I. et al 2002, Zhang D.S. 2002, Zhang D.S. et al 2004, Ritendra D. et al 2008 等）。

通常，在基于形状的图像检索中有三个问题需要解决（章毓晋 2003）。

（1）形状通常与图像中的特定目标对象有关，包含一定的语义信息，因此形状特征是比较颜色特征和纹理特征更高一层的特征。

（2）对形状的表达比对颜色和纹理的表达从本质上要复杂得多，要获得目标的形状参数，首先要进行图像分割，所以形状特征的提取会受到图像分割效果的影响，在没有相关领域的知识下，自动的分割方法很难将相应的目标区域准确的提取出来。同时，图像分割所得到的结果通常是具有一致性质的区域，而用户感兴趣的对象通常是具有现实意义的目标，要将不同区域合成为具有意义的目标，需要有高层知识的辅助或者人工干预。而且，描述目标的形状也是一个非常复杂的问题，人对形状的感觉是视网膜感受和现实世界的知识相结合的结果。事实上，目前还没有找到与人的主观感觉相一致的形状模型定义。除此之外，形状度量方法仍不具有很好的形状区分能力，不能有效的表达形状之间的相似性。

（3）从不同视角获得的图像中的目标形状可能会有很大差别，为了准确进行形状匹配，需要保证所提取的形状特征不受图像平移、缩放、旋转等变换的影响，这在许多基于形状的图像检索应用中起着关键作用。

以上存在的问题已经引起了研究者的关注，在一些探索性的研究中尝试利用数学形态学、神经网络、遗传算法等进行基于形状的图像检索，同时与此紧密相关的图像边缘检测及分割算法的研究也在进一步发展。

基于形状的检索依赖于对形状的表达和描述，因此如何描述形状是一个关键的问题。

按照形状表达的形式划分, 形状的描述符可分为两大类: 第一类是表达形状的目标区域边界轮廓的像素集合, 称为基于轮廓的形状描述符 (Contour-based); 第二类是表达形状的目标区域内的所有的像素集合, 称为基于区域的形状描述符 (Region-based) (Zhang D.S. 2002)。在每一类中, 根据描述符表达的是局部特征还是全局特征, 不同的描述方法还可进一步划分为基于结构 (structural) 的方法以及基于全局 (global) 的方法。详细的形状特征描述划分方法如图 4.1 所示 (Zhang D.S. et al 2004)。在形状特征的描述中, 还有许多简单的形状描述符, 如 *area*, *circularity*, *eccentricity*, *major axis orientation*, *bending energy*, *convexity*, *ratio of principle axis*, *circular variance* 及 *elliptic variance* 等 (Alajlan N. et al 1995, Yong I. et al 1974, Peura M. et al 1997, 章毓晋 1999, 周明全等 2007), 虽然这些特征仅仅是对形状特征的简单描述, 但它们却在实际中得到了广泛的应用。

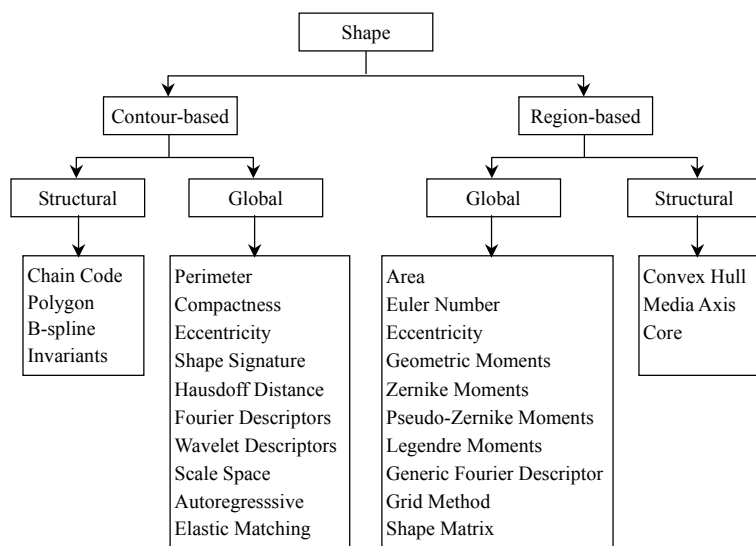


图 4.1 形状描述技术分类 (Zhang D.S. et al 2004)

## 4.2 基于轮廓的描述方法

基于轮廓的描述方法仅仅提取形状的轮廓信息, 这类描述方法一般有两种形式, 连续型 (即全局型) 和离散型 (即结构型)。连续型的描述方法不对轮廓进行分段处理, 往往是从全局轮廓提取特征矢量; 离散型的方法往往首先将轮廓划分为很多片段, 然后提取相应的特征。基于轮廓的常用方法有Freeman链码 (Freeman H. 1961)、*hausdorff distance* (Chetverikov D. et al 1999, Meier T. 1998, Rucklidge W.J. 1997, Scassellati B. et al 1994)、*shape contexts* (Belongie S. et al 2001)、曲率尺度空间 (Mokhtarian and Mackworth 1992)、BAS (Beam angle statistics) 描述符 (Arica N. et al 2003)、TAR (Triangel-Area Representation) 描述符 (Alajlan N. et al 2007)、傅里叶描述符 (Zhang D. S. 2002)、小波描述符 (杨翔英等 1999)、边界矩 (Chen, C.C. 1993, Gupta et al 1987)、基于内角的方法 (庄越挺 1998) 等。为了比较不同轮廓描述符的性能, 文献 (Veltkamp, R.C. et al 2006) 采用MPEG-7 CE-Shape-1 图像库 (Bober M. et al 1999) 对 11 种轮廓描述符的检索

效果进行了比较, 比较结果可参见<http://give-lab.cs.uu.nl/sidestep/>。

本节主要介绍常用的四种基于轮廓的形状描述方法: 链码、傅里叶描述符、曲率尺度空间描述符及小波描述符。

### 4.2.1 链码

链码 (Freeman H. 1961) 是对图像边界点的一种编码表示方法, 其特点是利用一系列具有特定长度和方向的相连直线段来表示目标的边界。因为每个线段的长度固定而方向数目取为有限, 所以只有边界的起点需用坐标表示, 其余点都只需用连续方向来代表偏移量。常用的链码有 4-方向链码及 8-方向链码, 其方向定义如图 4.2 所示。在此基础上, Freeman (Freeman H. 1974) 推广了原来的定义获得了广义链码, 并且还利用链码来抽签关键点, 从而生成一种相对于平移、旋转、尺度不变的表示方法 (Freeman H. 1978), 他还总结了链码的各种算法 (Freeman H. 1980)。由于链码表示简单并且所需的存储空间较小, 因此链码广泛的应用到形状编码及模式识别中 (Salem A. et al. 2005, Liu Y.K. et al 2005)。同时, 在链码定义的启发下, 还出现了其他一些定义链码的方法, 如夹角链码 (赵宇等 2004) 及可变夹角链码 (刘淑娟 2005)、3OT 链码 (Bribiesca E. 2000)、VCC 链码 (Bribiesca E. 1999) 等, 文献 (Hermilo S.-M. et al 2007) 对多种链码描述轮廓的性能进行了比较分析。

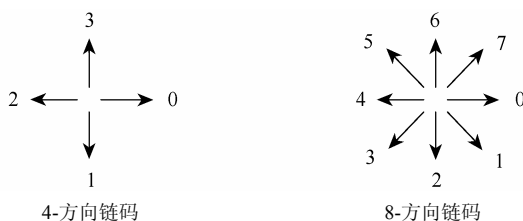


图 4.2 4-方向链码及 8-方向链码

针对方向链码, 起点的变化、尺度的变化、目标的旋转等都会引起链码串的变化, 而在对形状进行描述时, 关键就是要使所选择的描述符满足尺度、旋转和平移不变性。因此, 往往需要进行起点归一化及旋转归一化 (章毓晋 1999)。本节仅简单的介绍链码的基本概念, 有关链码在图像检索技术中的应用, 将在 4.5 节详细讨论。

### 4.2.2 傅里叶形状描述符

傅里叶形状描述符 (Fourier Shape Descriptors) 是一种广泛应用的形状描述符, 其基本思想是用物体边界的傅里叶变换作为其形状描述 (Chellappa R. et al 1984, Arbter K. et al 1990, Granlund G. 1972, Kauppinen H. et al 1995, Krzyzak A. et al 1989, Zhang D.S. et al 2002a, Zhang D.S. et al 2002b, 庄越挺等 2002, Zhang D.S. 2002)。假设一个二维物体的轮廓是由一系列坐标为  $(x_s, y_s)$  的像素组成, 其中  $0 \leq s \leq N-1$ , 而  $N$  是轮廓上像素的总数。从这些边界点的坐标中可以推导出四种形状表达, 分别是曲率函数 (curvature function)、质心距离 (centroid distance)、复坐标函数 (complex coordinates function) 和弦长函数 (chord function) (Zhang D.S. 2002)。文献 (庄越挺等 2002) 给出了对傅里叶描述符的描述。



轮廓线上某一点的曲率定义为轮廓切向角度相对于弧长的变化率。曲率函数  $K(s)$  可以表示为

$$K(s) = \frac{d}{ds} \theta(s) \quad (4-1)$$

式中,  $\theta(s)$  是轮廓线的切向角度, 定义为

$$\left. \begin{aligned} \theta(s) &= \arctan \left( \frac{y'_s}{x'_s} \right) \\ y'_s &= \frac{dy_s}{ds} \\ x'_s &= \frac{dx_s}{ds} \end{aligned} \right\} \quad (4-2)$$

质心距离定义为从物体边界点到物体中心  $(x_c, y_c)$  的距离为

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (4-3)$$

复坐标函数是用复数所表示的像素坐标为

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (4-4)$$

对这种复坐标函数的傅里叶变换会产生一系列复数系数。这些系数在频率上表示了物体形状, 其中低频分量表示形状的宏观属性, 高频分量表达了形状的细节特征。形状描述符可以从这些变换参数中得出。为了保持旋转无关性, 可以仅仅保留参数的大小信息, 而省去相位信息。缩放的无关性是通过将参数的大小除以直流分量 (DC) (即第一个非零参数) 的大小来保证的。另外, 变换无关性是基于轮廓的形状表示所固有的特点。

对于曲率函数和质心距离函数, 我们只考虑正频率的坐标轴, 因为这时函数的傅里叶变换是对称的, 即有  $|F_{-i}| = |F_i|$ 。基于曲率函数的形状描述符表示为

$$f_K = [ |F_1|, |F_2|, \dots, |F_{M/2}| ] \quad (4-5)$$

式中,  $F_i$  表示傅里叶变换参数的第  $i$  个分量。类似的, 由质心距离所导出的形状描述符为

$$f_R = \left[ \frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{M/2}|}{|F_0|} \right] \quad (4-6)$$

对于复坐标函数, 正频率分量和负频率分量被同时采用。由于  $DC$  参数与形状的所处的位置有关而被省去。因此, 第一个非零的频率分量被用来对其他变换参数进行标准化。复坐标函数所导出的形状描述符为

$$f_Z = \left[ \frac{|F_{-(M/2-1)}|}{|F_1|}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{M/2}|}{|F_1|} \right] \quad (4-7)$$

为保证数据库中所有物体的形状特征都有相同的长度, 在实施傅里叶变换之前需要将所有边界点的数目统一, 这样就可以采用快速傅里叶变换来提高算法效率。

### 4.2.3 曲率尺度空间描述符

曲率尺度空间描述符 (Curvature Scale Space Descriptor, CSSD) 是由 Mokhtarian 等引入 (Mokhtarian F. et al 1992), 并成功应用于形状特征的描述及检索中 (Mokhtarian F.

et al 1996a, Mokhtarian F. et al 1996b, Abbasi S. et al 1999, Abbassi S. et al 2000)。目前, 该描述符成为 MPEG-7 标准中轮廓描述的一种方法。该方法的基本思想是: 基于人们在认知物体时, 倾向于将物体的形状分解成凹和凸的部分来比较这一特性 (Manjunath B.S. et al 2002)。CSSD 是在多尺度下根据变形点将轮廓“分解”成凹和凸部分, 这些变形点定义为轮廓曲线上曲率为零的点。

给出物体的闭合轮廓曲线, 对曲线的两直角坐标  $x$  及  $y$  进行参数化, 采用自然参数即曲线的弧长为参数, 以任意一点为起点, 顺时针跟踪轮廓, 并归一化使得弧长参数  $l \in [0, 1]$ , 在曲线起始点处  $l=0$ , 终点处  $l=1$ , 轮廓曲线表示为:  $C = \{x(l), y(l)\}$ ,  $l \in [0, 1]$ 。轮廓若为闭合, 则起点和终点是重合的, 且  $x(l)$  和  $y(l)$  为以 1 为周期的周期函数。曲率计算式如:

$$k(l) = \frac{x'(l)y''(l) - x''(l)y'(l)}{((x'(l))^2 + (y'(l))^2)^{\frac{3}{2}}} \quad (4-8)$$

式中  $x'(l)$ 、 $y'(l)$ 、 $y''(l)$  及  $x''(l)$  分别代表一阶及二阶导数。

为了获得不同分辨率下轮廓的曲率, 需要对轮廓进行演化。用核宽为  $\sigma$  的一维高斯函数  $g(l, \sigma)$  进行卷积平滑, 然后在平滑的基础上再次计算曲率。图 4.3 给出了一轮廓在不同尺度下的曲率过零点的情况, 图 4.4 给出了轮廓的 CSSD 的示意图 (Mokhtarian F. et al 1996a, Mokhtarian F. et al 1996b)。

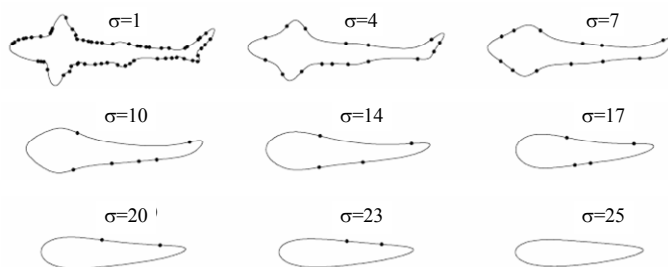


图 4.3 曲率过零点

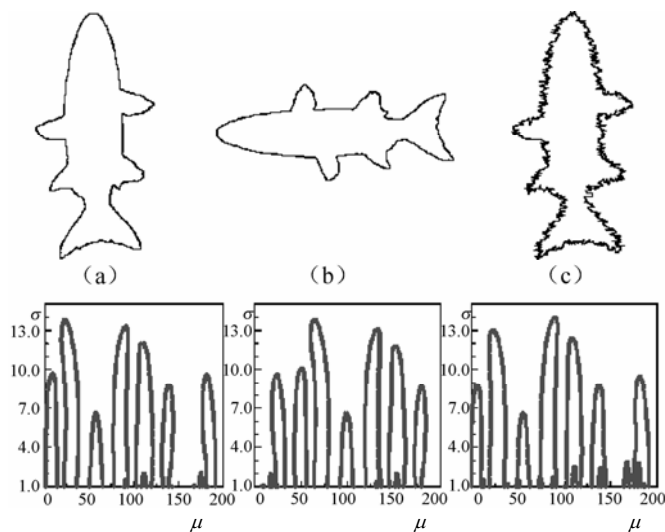


图 4.4 CSSD 示意图

#### 4.2.4 小波描述符

利用小波变换对图像轮廓进行描述（杨翔英等 1999, Ohm J.R. et al 2000, Tieng Q.M. et al 1997），首先要定义小波函数族。小波函数族定义为（崔锦泰著 程正兴译 1997）

$$\psi_{mn}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (4-9)$$

假设图像的轮廓函数为  $f(t)$ ，它的小波变换系数为

$$c_{mn}(t) = \int_{-\infty}^{\infty} f(t) \psi_{mn}(t) dt \quad (4-10)$$

利用小波系数可以重建  $f(t)$ ，重建过程为

$$f(t) = \sum_{m=m_0+1}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} \psi_{mn}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} c_{mn} \psi_{mn}(t) \quad (4-11)$$

式中， $m_0$  与截断系数时所需的精度相关。

假设尺度函数为  $S_{mn}(t) = 2^{-m/2} S(2^{-m}t - n)$ ，结合小波重建公式，则有

$$f(t) = \sum_{n=-\infty}^{\infty} c_{mn} S_{mn}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} c_{mn} S_{mn}(t) \quad (4-12)$$

如果  $c_{mn} S_{mn}(t)$  称为尺度系数，则  $c_{mn} \psi_{mn}(t)$  称为小波系数，所有的小波系数组成轮廓相对应的小波轮廓描述符。

小波轮廓描述符在粗尺度上给出了形状的全局信息，在细尺度上给出了局部信息。由于小波变换提供了多分辨率表示，因此识别过程可以根据输入图像或目标而灵活调整。但小波描述符依赖于目标曲线的起始点，也就是说，同一目标的两条不同采样曲线的小波表示可能因为起始点不同而有很大的差异（Loncrie S. 1998, 丁险峰等 2001）。

### 4.3 基于区域的描述方法

基于区域的表示方法将区域形状当作一个整体来看待，该方法有效地利用了区域内的所有像素，因而受噪声和形状变化的影响相对较小。同基于轮廓的形状表示方法一样，基于区域的形状表示方法也分为全局型和局部型两种类型（Zhang D.S. et al 2004）。形状的区域特征主要有区域的面积、欧拉数、离散度、偏心率、区域不变矩、区域骨架、几何不变矩、Legendre 矩、Zernike 矩、伪 Zernike 矩、旋转矩、复数矩、通用傅里叶描述符、角半径变换等方法。这里主要介绍几何不变矩（Geometric Moment Invariants）、Zernike 矩、角半径变换及通用傅里叶描述符四种常用的基于区域的形状描述方法。

#### 4.3.1 几何不变矩

图像的矩函数在模式识别、目标分类中得到了广泛的应用。1962 年，Hu（Hu M.K. 1962）提出了图像识别的不变矩理论，并且首次提出了基于代数不变量的矩不变量，并通过对几何矩的非线性组合，导出了一组对于图像平移、旋转和尺度变化不变的矩。不变矩是图像的一种统计特征，它利用图像灰度分布的各阶矩来描述图像灰度的分布特性。

对于离散的数字图像  $f(x, y)$  的  $p+q$  阶矩定义为（章毓晋 1999）

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (4-13)$$

其  $p+q$  阶中心矩定义为

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (4-14)$$

式中,  $\bar{x} = m_{10}/m_{00}$ ,  $\bar{y} = m_{01}/m_{00}$  表示图像的区域重心。中心矩表示了图像内不同灰度级的像素相对于其重心是如何分布的, 因此中心矩具有位置无关性。为了获取针对图像缩放无关的性质, 可以对该中心矩进行规格化操作, 规格化后的中心矩表示为

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (4-15)$$

式中,  $\gamma = \frac{p+q}{2} + 1, p+q = 2, 3, \dots$  中心规格矩对于物体缩放、平移和旋转均保持不变 (夏德深等 1997)。基于规格化的二阶和三阶中心矩, 可以导出如下 7 个矩组:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (4-16)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (4-17)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (4-18)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (4-19)$$

$$\begin{aligned} \phi_5 = & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[ (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] + \\ & (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[ 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \end{aligned} \quad (4-20)$$

$$\phi_6 = (\eta_{20} - \eta_{02}) \left[ (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (4-21)$$

$$\begin{aligned} \phi_7 = & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \left[ (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] + \\ & (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[ 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \end{aligned} \quad (4-22)$$

上述 7 个不变矩组称为 Hu 不变矩。Hu M.K. 证明了这个矩组中的  $\phi_1 \sim \phi_6$  具有尺度、平移和旋转不变性; 而  $\phi_7$  只具有尺度和平移不变性, 不具有旋转不变性, 仅在镜面对称时保持不变 (Liao S. X. et al 1996)。

在 Hu 不变矩的基础上, 刘进等 (刘进等 2004) 提出了一种快速有效的推导不变矩的方法, 并将图像的 7 个不变矩进行了推广, 导出了 5 个新的不变矩表达式  $\phi_8 \sim \phi_{12}$ , 其定义如下:

$$\phi_8 = 2 \left\{ \eta_{11} \left[ (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] - (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \right\} \quad (4-23)$$

$$\begin{aligned} \phi_9 = & [(\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12}) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})](\eta_{20} - \eta_{02}) + \\ & 2\eta_{11}[(3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})] \end{aligned} \quad (4-24)$$

$$\begin{aligned} \phi_{10} = & [(3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})](\eta_{20} - \eta_{02}) - \\ & 2\eta_{11}[(\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})] \end{aligned} \quad (4-25)$$

$$\phi_{11} = (\eta_{04} + \eta_{40} - 6\eta_{22})^2 + 16(\eta_{31} - \eta_{13})^2 \quad (4-26)$$

$$\phi_{12} = (\eta_{04} + \eta_{40} - 6\eta_{22}) \left[ (\eta_{20} - \eta_{02})^2 - 4\eta_{11}^2 \right] + 16\eta_{11}(\eta_{31} - \eta_{13})(\eta_{20} - \eta_{02}) \quad (4-27)$$

同时证明了所导出的矩  $\phi_8$ 、 $\phi_9$ 、 $\phi_{11}$ 、 $\phi_{12}$  具有比经典的  $\phi_5$  和  $\phi_7$  更好的特性。

### 4.3.2 Zernike矩

Zernike 矩定义为 (Teague M.R. 1980)

$$Z_{nm} = \frac{n+1}{\pi} \sum_y \sum_x V_{nm}^* f(x, y), x^2 + y^2 \leq 1 \quad (4-28)$$

式中,  $V_{nm}^*(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp(jm\theta)$ 。  $R_{nm}(\rho)$  的定义为

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!((n+|m|)/2-s)!((n-|m|)/2-s)!} \times \rho^{n-2s} \quad (4-29)$$

式中,  $n$  和  $m$  为非负整数, 并且必须满足  $n-|m|$  为偶数且  $n \geq |m|$ 。

自从 Zernike 矩的概念被引入以来, Zernike 矩便因其具有优良的旋转不变性而在模式识别等领域得到广泛的应用 (Kim Y.-S. et al 1998, Kim W.-Y. et al 2000)。在图像分析中, 由于 Zernike 多项式的正交性, 可以使信息冗余达到最优, 它的递归性质使得矩的快速算法成为可能。Zernike 矩还有一个重要的特性: 图像旋转一定角度后的 Zernike 矩与原图像的 Zernike 矩有非常简单的关系, 就是通过旋转 Zernike 矩仅仅相位发生变化而幅值保持不变。

虽然 Zernike 矩具有旋转不变性, 但却不具备尺度不变性, 因此在采用 Zernike 矩描述形状时, 首先还需要进行归一化处理。

矩是区域形状的一种常用描述方法, 除了上述的几何不变矩及 Zernike 矩之外, 还有其他很多区域矩的描述方法, 有关矩的描述基本上分为两大类型: 正交矩及非正交矩。常用的正交矩有: Legendre Moments, Zernike Moments, pseudo-Zernike Moments 等; 常用的非正交矩有: Geometric Moments (即几何不变矩)、Complex Moments、Rotation Moments 等 (Zhang D.S. et al 2004)。文献 (Teh C.-H. 1988, Liao S.X. 1996) 对上述不同类型的矩的特性进行了比较分析及改进。在上述的区域形状的矩描述符中, Zernike 矩的性能是最优的 (Zhang D.S. et al 2004), 且 MPEG-7 标准中已经将 Zernike 矩列为一种标准的区域描述符 (Jeannin S. (Ed.) 2000)。

### 4.3.3 ART ( Angular Radial Transformation )

ART 是 MPEG-7 推荐的另一个基于区域的形状描述符 (Manjunath B.S. et al 2000), 也是一种基于矩的图像描述符。它使用一组角半径变换系数, 描述单个连通区域或者多个不连通区域, 并且对旋转具有鲁棒性。角半径变换是定义在极坐标下的一个单位圆内的二维复变换, 是一种正交变换, 对噪声具有鲁棒性 (Bober M. 2001)。ART 变换的系数定义为

$$F_{nm} = \langle V_{nm}(\rho, \theta) f(\rho, \theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta \quad (4-30)$$

式中,  $F_{nm}$  是在序数  $(n, m)$  的 ART 系数,  $f(\rho, \theta)$  是在极坐标上的一个图像灰度方程, 是 ART 的基本函数, 如图 4.5 所示。 $V_{nm}^*(\rho, \theta)$  为其复共轭, 角半径变换系数的基本函数沿着角方向和放射方向是可分离的, 它的表达式为

$$V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho) \quad (4-31)$$

式中,  $A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta)$ ,  $R_n(\rho) = \begin{cases} 1, n=0 \\ 2 \cos(\pi n \rho), n \neq 0 \end{cases}$ 。

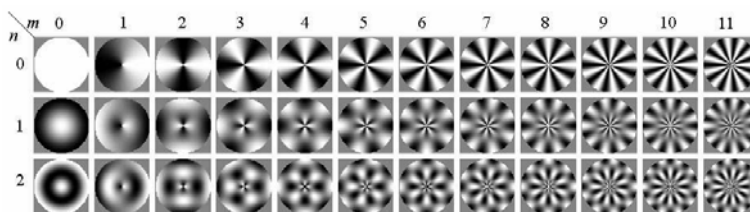


图 4.5 ART 基本变换

### 4.3.4 通用傅里叶描述符

通用傅里叶描述符 (Generic Fourier Descriptor, GFD) (Zhang D.S. et al 2002, Zhang D.S. 2002) 采用了修正的平面极坐标傅里叶变换, 对图像进行采样, 将采样的信息重新绘制在笛卡儿直角坐标系下, 再对该直角坐标系下的图像做傅里叶变换, 如图 4.6 所示。

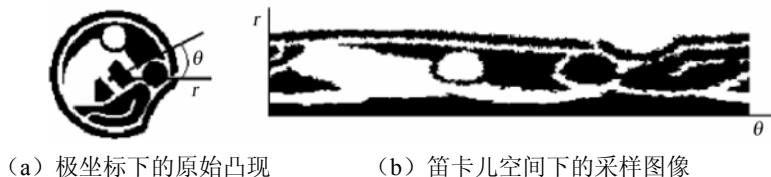


图 4.6 通用傅里叶描述符示意图

修正的极坐标傅里叶变换定义如下:

$$GF(\rho, \theta) = \sum_r \sum_i f(r, \theta_i) \exp \left[ j2\pi \left( \frac{r}{R} \rho + \frac{2\pi i}{T} \phi \right) \right] \quad (4-32)$$

式中,  $0 \leq r = \left[ (x - x_c)^2 + (y - y_c)^2 \right]^{\frac{1}{2}} \leq R$ ,  $\theta_i = i(2\pi/T) (0 \leq i \leq T)$ ,  $(x_c, y_c)$  是形状的中心点,  $R, T$  分别为半径和角度的分辨率。GFD 具有平移不变性, 为了获取尺度和旋转不变性, 可进行规范化

$$GFD = \left\{ \frac{GF(0,0)}{\text{area}}, \frac{GF(0,1)}{GF(0,0)}, \dots, \frac{GF(0,n)}{GF(0,0)}, \dots, \frac{GF(m,0)}{GF(0,0)}, \dots, \frac{GF(m,n)}{GF(0,0)} \right\} \quad (4-33)$$

式中, area 表示极坐标采样时形状所占的最大面积,  $m$  是最大半径频数,  $n$  是最大半径角度频数。

## 4.4 简单几何参数描述符

当图像中的目标个数较少 (大多数只有一个目标) 时, 常用描述图像目标的几何参数作为形状特征。常用的几何参数描述符也可分为基于轮廓的方法及基于区域的方法 (章毓晋 1999)。

### 4.4.1 基于轮廓的方法

#### (1) 边界的长度

边界长度是一种简单的边界全局特征, 它是边界所包围区域的轮廓的周长。边界长度一般常用 4-方向连通边界或 8-方向连通边界表示, 并相应的得到一个近似长度。

#### (2) 边界的直径

它是边界上相隔最远的两点之间的距离, 即这两点之间的直连线段长度。有时这条直线也称为边界的主轴或长轴 (与此垂直且最长的与边界的两个交点间的线段也称边界的短轴), 其长度和取向对描述边界都很有用。

#### (3) 曲率

曲率是斜率的改变率, 它描述了边界上各点沿边界方向变化的情况。对于离散的轮廓

图来说, 曲率往往会受到噪声及边缘细节的影响。

#### (4) 形状数

形状数是基于链码的一种边界形状描述符, 形状数是值最小的(链码)差分码, 具体计算方法可参阅参考文献(章毓晋 1999)。形状数提供了一种有效的形状度量方法, 它的每个阶是唯一的, 不随边界的旋转和尺度的变化而改变; 另外形状数还提供了一种形状相似性的度量方法。

#### (5) 矩描述

目标的边界可看做由一系列曲线段组成, 矩描述的基本思想是: 将任意给定的曲线段表示成一个 1-D 函数  $f(r)$  ( $r$  为任意变量), 进一步可把  $f(r)$  的线下面积归一化成单位面积并把它看成一个直方图, 则  $r$  变成一个随机变量。可用曲线段的矩来描述曲线段, 而进一步描述整个边界。

### 4.4.2 基于区域的方法

区域描述方法主要包括以下参数。

#### (1) 区域面积

区域面积是区域的一个基本特征, 它描述区域的大小。设正方形像素的边长为单位长, 则计算区域面积就是对属于区域的像素计数。

#### (2) 区域重心

区域重心是一种全局描述符, 区域重心的坐标是根据所有属于区域的点计算出来的, 其计算方式如下:

$$\begin{aligned}\bar{x} &= \frac{1}{A} \sum_{(x,y) \in R} x \\ \bar{y} &= \frac{1}{A} \sum_{(x,y) \in R} y\end{aligned}\tag{4-34}$$

式中,  $R$  表示某一区域,  $A$  表示区域面积(即区域像素数)。

#### (3) 区域灰度(密度)

目标的灰度特性要结合原始灰度图和分割图来得到。常用的区域灰度特征有目标灰度的最大值、最小值、中值、平均值、方差以及高阶矩等统计量, 它们多可借助灰度直方图得到。

#### (4) 欧拉数

欧拉数是一种拓扑描述符, 区域的拓扑性质对区域的全局描述很有用, 这种性质既不依赖距离, 也不依赖基于距离测量的其他特性。设  $H$  表示区域内的空数,  $C$  表示区域内的连通组元的个数, 则欧拉数可定义如下:

$$E = C - H\tag{4-35}$$

#### (5) 形状参数

形状参数是根据区域的周长和区域的面积计算出来的, 其计算方式如下:

$$F = \frac{\|B\|^2}{4\pi A}\tag{4-36}$$

式中,  $B$  为区域的周长。可以看出, 圆形的形状参数为 1, 当区域为其他形状时, 形状参数大于 1。形状参数在一定程度上描述了区域的紧凑性(compactness), 且对尺度变换及

旋转变换不敏感。

#### (6) 偏心率

偏心率 (Eccentricity) 也可叫做伸长度 (Elongation), 它也在一定程度上描述了区域的紧凑性。由惯性推出的偏心率计算公式为

$$E = \sqrt{\frac{(A+B) - \sqrt{(A-B)^2 + 4H^2}}{(A+B) + \sqrt{(A-B)^2 + 4H^2}}} \quad (4-37)$$

式中,  $A = \sum m_i(y_i^2 + z_i^2)$ ,  $B = \sum m_i(x_i^2 + z_i^2)$ ,  $C = \sum m_i(x_i^2 + y_i^2)$  分别是刚体绕  $x$ ,  $y$ ,  $z$  轴的转动惯量, 而  $F = \sum m_i y_i z_i$ ,  $G = \sum m_i z_i x_i$ ,  $H = \sum m_i x_i y_i$  均称作惯性积。

偏心率不受区域平移、旋转和尺度变换的影响。

#### (7) 球状性

二维区域的球状性 (Sphericity) 的计算需要用到区域的重心, 它定义为

$$S = \frac{r_i}{r_c} \quad (4-38)$$

式中,  $r_i$  代表区域内切圆的半径, 而  $r_c$  代表区域外接圆的半径, 两个圆的圆心都在区域的重心上。当区域为圆时球状性的值达到最大为 1, 而当区域为其他形状时则小于 1。它也不受平移、旋转和尺度变化的影响。

#### (8) 圆形性

圆形性 (Circularity) 是用区域的所有边界点定义的特征量, 其计算方式为

$$C = \frac{\mu_R}{\sigma_R} \quad (4-39)$$

式中,  $\mu_R = \frac{1}{K} \sum_{k=0}^{K-1} \|(x_k, y_k) - (\bar{x}, \bar{y})\|$  为从区域重心到边界点的平均距离,  $\sigma_R = \frac{1}{K} \sum_{k=0}^{K-1} \left[ \|(x_k, y_k) - (\bar{x}, \bar{y})\| - \mu_R \right]^2$  为从区域重心到边界点的距离的均方差,  $K$  为边界像素点的个数。当区域趋向圆形时, 圆形性是单增趋向无穷的, 同时圆形性也不受区域的平移、旋转和尺度变化的影响。

## 4.5 基于状态矩阵描述方法

### 4.5.1 状态矩阵定义

为了描述图像的形状信息, 可将图像的色调、饱和度和亮度三个灰度分量分别划分为  $n \times n$  个小区块子块, 对于某一区域子块, 如果该区域内像素间的灰度值差别较小, 在人眼看来, 该区域表现的就比较平坦; 如果该区域内像素间的灰度值差别较大, 则该区域表现得就比较起伏。依据这种特性, 我们可以将图像的区域子块划分为不同的类型。由于区域内像素灰度值的变化可通过该区域内像素灰度值的标准差来反映, 对于某一子块, 如果其内部像素灰度值的标准差很小, 则表明该子块内像素的灰度值比较接近, 因此该子块对应图像中灰度变化平坦的区域; 如果其内部像素灰度值的标准差很大, 则表明该子块内像素的灰度值变化较大, 而只有包含图像边缘细节的区域子块才具有这种特性; 介于这二者



之间的子块，像素间的灰度值相差则不是很大，这种子块在一定程度上反映了图像纹理特征。依据子块内像素灰度值的标准差，将图像区域子块划分为平坦态、纹理态和边缘态 3 种状态（孙君顶等 2004a）。

设  $\sigma$  表示子块内像素灰度值的标准差， $\sigma$  定义如下：

$$\sigma = \left( \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m (f(i, j) - \mu)^2 \right)^{\frac{1}{2}} \quad (4-40)$$

式中， $\mu = \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m f(i, j)$ ， $f(i, j)$  表示坐标  $(i, j)$  处像素的灰度值， $l \times m$  表示所划分子块的大小。

设  $\alpha_1$ 、 $\alpha_2$  ( $0 < \alpha_1 < \alpha_2$ ) 为事先给定的阈值，则有如下图像分块状态的定义：

**定义 1：** 平坦态 对于图像中的某一分块，如果满足条件  $\sigma < \alpha_1$ ，则定义该子块的状态为“平坦态”，并用“0”表示该种状态。

**定义 2：** 边缘态 对于图像中的某一分块，如果满足条件  $\sigma > \alpha_2$ ，则定义该子块的状态为“边缘态”，并用“2”表示该种状态。

**定义 3：** 纹理态 对于图像中的某一分块，如果满足条件  $\alpha_1 < \sigma < \alpha_2$ ，则定义该子块的状态为“纹理态”，并用“1”表示该种状态。

对于如何确定参数  $\alpha_1$  和  $\alpha_2$  的取值，在 10 000 幅的图像库中任意选取 200 幅图像，所选取的图像包括风景、人、交通工具、动物、花卉、建筑等多个类别，这样可以使统计的结果具有一般性，而不会仅仅局限于某一类别的图像，同时，所选取的图像被划分为  $32 \times 32$  个分块。对子块内像素灰度值标准差的分布进行统计，统计结果如图 4.7 所示。其中横坐标表示标准差的变化情况，纵坐标表示在 200 幅图像中不同标准差取值所对应的分块总数。

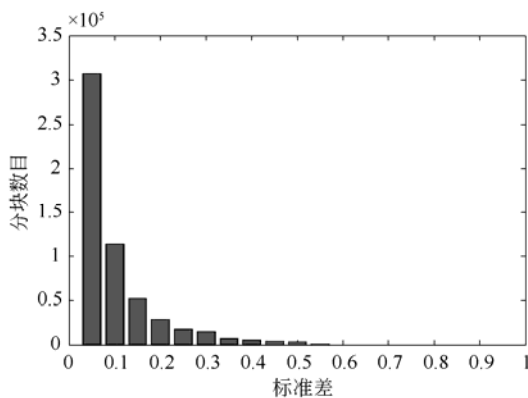


图 4.7 分块灰度标准差分布

由图 4.7 可以看出图像子块内像素灰度值的标准差主要分布在区间  $[0, 0.6]$ ，且分布在区间  $[0, 0.05]$  的子块占了大多数，而分布在区间  $[0.6, 1]$  的子块几乎没有，根据这种统计结果及基于图像整体特性的考虑，在对图像分块状态的划分时，取  $\alpha_1 = 0.05$ 、 $\alpha_2 = 0.4$ 。

根据上述对子块类型的定义，可以将图像中的每一区域子块分别采用其状态值来表示，这种采用分块的状态来表示的图像矩阵称为图像的状态矩阵。这样，与彩色图像对应

的每个灰度图像均可转化为由 3 种状态  $\{0, 1, 2\}$  构成的二维状态矩阵, 状态矩阵中状态的变化反映了图像的形状变化。

图像的状态矩阵反映了图像形状的变化, 并且将图像的形状变化映射到图像分块状态的交替上, 这大大减少了进行形状特征提取的计算量。如何由图像的状态矩阵提取图像的形状变化信息是本节的关键问题, 为此我们提出采用基于马尔可夫链(孙君顶等 2004)和状态相关图(State Correlogram)(孙君顶等 2005)两种方法来分别从图像的状态矩阵中提取图像的形状变化信息。

## 4.5.2 基于马尔可夫链形状特征提取

### 1. 马尔可夫链

马尔可夫链(马氏链)最初是由 Markov 提出和研究的一类随机过程, 目前已广泛地应用于计算机、通信、自动控制等各个方面。设离散参数  $N_0 = \{0, 1, 2, \dots\}$ , 状态空间  $S = \{1, 2, \dots\}$ , 则有如下定义(林元烈 2002)。

**定义 4:** 如果随机序列  $\mathbf{X} = \{\mathbf{X}_c, c \geq 0\}$  对于任意的  $a_0, a_1, \dots, a_c, a_{c+1} \in S, c \in N_0$  及概率  $p\{\mathbf{X}_0 = a_0, \mathbf{X}_1 = a_1, \dots, \mathbf{X}_c = a_c\} > 0$ , 有

$$p\{\mathbf{X}_{c+1} = a_{c+1} | \mathbf{X}_0 = a_0, \mathbf{X}_1 = a_1, \dots, \mathbf{X}_c = a_c\} = p\{\mathbf{X}_{c+1} = a_{c+1} | \mathbf{X}_c = a_c\} \quad (4-41)$$

则称该随机序列为马尔可夫链。式(4-41)刻画了马尔可夫链的特性, 称为马尔可夫性(或无后效性)。

**定义 5:** 系统在  $c$  时刻处于状态  $a_i$  的条件下经过  $k$  步转移, 与  $c+k$  时转移到状态  $a_j$  的条件概率  $p_{ij}^{(k)}(c)$  称为  $\mathbf{X}$  在  $c$  时刻的  $k$  步转移概率; 第  $(i, j)$  元素为  $p_{ij}^{(k)}(c)$  的矩阵  $\mathbf{P}^{(k)}(c)$  称为  $\mathbf{X}$  在  $c$  时刻的  $k$  步转移概率矩阵。特别地,  $k=1$  (在  $c$  时)的一步转移概率和一步转移概率矩阵分别简记为  $p_{ij}(c)$  和  $\mathbf{P}(c)$ 。

在马氏链中,  $k$  步转移概率和一步转移概率之间是等价的, 即可相互表示。对于马氏链有下述定理存在。

**定理 1:**

$$p_{ij}^{(k+m)}(c) = \sum_l p_{il}^{(k)}(c) p_{lj}^{(m)}(c+k), \quad n, k, m \geq 0, i, j \in N_0 \quad (4-42)$$

$$\mathbf{P}^{(k+m)}(c) = \mathbf{P}^{(k)}(c) \mathbf{P}^{(m)}(c+k), \quad c, k, m \geq 0 \quad (4-43)$$

其中, 式(4-42)是式(4-43)的矩阵形式。上述等式称为 Chapman-Kolmogorov (简称 C-K) 方程。

C-K 方程得到直观意义如下: 系统在  $c$  时从状态  $a_i$  出发, 经过  $k+m$  步于  $c+k+m$  时转移到状态  $a_j$ , 可以先在  $c$  时从状态  $a_i$  出发, 经过  $k$  步于  $c+k$  时转移到某个中间状态  $a_l$ , 再在  $c+k$  时从中间状态  $a_l$  出发, 经过  $m$  步于  $c+k+m$  时转移到状态  $a_j$ , 而中间状态  $a_l$  取遍全状态空间  $S$ 。且对于马氏链  $\mathbf{X}$ , 其  $k$  步转移概率由一步转移概率完全确定。

同时, 由文献(林元烈 2002)的分析可知, 一个马尔可夫链运动规律的概率特性取决于它的转移概率矩阵特性。因此, 本节在对图像的状态矩阵进行处理时, 即利用了马尔可夫链的这种特性。

## 2. 特征提取与表示

众所周知,在二维图像空间中,像素的空间位置越近,则其相关性就越强,因而对于图像的区域子块,空间位置越近,相应的相关性也越强。同样,对于图像的状态矩阵,也具有相同的特性,即空间位置越近的状态之间的相关性越强。为了从图像的状态矩阵中提取图像的形状信息,我们采用Z字型扫描的方法将图像的状态矩阵转化为一维随机状态序列 $\{y_1, y_2, \dots, y_{n \times n}\}$  ( $y_e$ 表示子块 $e$ 的状态,  $e=1, 2, 3, \dots, n \times n$ ),该序列对应的状态空间为 $\mathbf{U} = \{0, 1, 2\}$ 。采用Z字型扫描方法可以尽可能将图像状态矩阵中相关性强的状态联系在一起,从而更有利于提取图像的形状特征。在图像的一维状态序列中,位置越靠近的两个状态的相关性越强,若某一状态为平坦态,则其相邻的两个状态在很大程度上可能为平坦态,随着不同状态在序列中距离的增大,它们间的相关性会逐步减弱。

针对图像一维状态序列,做如下假设:在图像的状态序列中,任何一种状态的出现仅与该状态的前一状态有关,而与该状态更前的状态及后继状态无关,即当前子块状态是“平坦态”、“纹理态”还是“边缘态”仅仅与它前一子块的状态有关。在此条件下,图像状态序列满足:

(1) 序列的状态数目有限(状态数为3);

(2) 对于某一幅图像,其对应的一维状态序列可唯一确定,且该状态序列同时间的变化没有关系;

(3) 进入某状态的概率仅与该状态之前的状态有关。

根据前述的马氏链的特性即C-K方程可知,对于马氏链 $\mathbf{X}$ ,其 $k$ 步转移概率由一步转移概率完全确定,因此这里仅考虑图像状态序列的一步转移概率矩阵的计算方法,针对多步转移概率矩阵,可通过一步转移概率矩阵求得。设 $r$ 及 $t$ 是图像状态随机序列中任意两个邻接状态,从而转移概率矩阵可表示为

$$\mathbf{P} = \{p_{r,t}\} = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{bmatrix} \quad (4-44)$$

式中,  $p_{r,t} = p\{y_e = t | y_{e-1} = r\} = \frac{N_{r,t}}{\sum_{z=0}^2 N_{r,z}}$ ,  $e=1, 2, 3, \dots, n \times n$ ,  $r, t \in \mathbf{U}, \mathbf{U} = \{0, 1, 2\}$ 。

若将上述的邻接状态 $r$ 及 $t$ 称为一状态对,记为 $(r, t)$ ,则式(4-44)中 $N_{r,t}$ 表示在图像随机状态序列中,状态对 $(r, t)$ 出现的总次数。依据上述方法,就可以分别计算出彩色图像3个灰度分量图像的转移概率矩阵。转移概率矩阵反映了图像状态矩阵的状态变化,这种状态变化又反映了图像的形状的变化,因而可将所求出的3个转移概率矩阵作为图像形状特征的描述。

对于由图像状态序列所求取的转移概率矩阵,同数学意义上由随机序列所求取的转移概率矩阵稍有不同,假设随机状态序列的状态空间 $\mathbf{U} = \{0, 1, 2\}$ ,则在数学意义上,转移概率矩阵 $\{p_{r,t}\}$ 满足

$$\sum_{t=0}^2 p_{r,t} = 1, r \in \mathbf{U} \quad (4-45)$$

但对于通过图像状态矩阵转化来的图像状态序列却并不完全满足式(4-45)。由于具

体到某幅图像不同灰度分量对应的随机状态序列，其状态空间为图像库状态空间的子集，因此，如果某图像对应的随机状态序列中不存在某种状态，那么采用式（4-44）来计算状态序列的转移概率矩阵时就会出现  $\sum_{t=0}^2 p_{r,t} = 0$  的情况，即分母为零的情况。为了防止这种情况，在实际处理过程中，可首先判断  $\sum_{t=0}^2 p_{r,t}$  的取值，如果  $\sum_{t=0}^2 p_{r,t} = 0$ ，则直接令  $p_{r,0} = p_{r,1} = p_{r,2} = 0$ ，以防止出现分母为零的情况。

#### 4.5.3 基于状态相关图的特征提取方法

为了有效地提取图像颜色的空间分布特征，Jing Huang (Huang J. 1998) 提出采用颜色相关图 (Color Correlogram) 的方法来描述不同颜色的像素按照空间距离在图像空间的分布特征，采用与 Color Correlogram 类似的方法，我们这里采用状态相关图 (State Correlogram) 来反映状态空间中不同状态按照空间距离的分布特征。

##### 1. 状态相关图

设  $\mathbf{T}$  表示图像某颜色通道的状态矩阵，其状态数记为  $L$ ， $b_1, b_2, \dots, b_L$  表示不同的状态值， $t(x, y)$  表示状态矩阵中任一状态点， $\mathbf{T}(t)$  表示该点的状态值。令  $\mathbf{T}_b \triangleq \{t | \mathbf{T}(t) = b\}$ ，则  $t \in \mathbf{T}_b$  等价于  $t \in \mathbf{T}$  且  $\mathbf{T}(t) = b$ 。对于状态矩阵中任意两个状态点  $t_1 = (x_1, y_1)$ ， $t_2 = (x_2, y_2)$ ，定义状态点  $t_1$  和  $t_2$  间的空间距离为

$$k = |t_1 - t_2| \triangleq \max\{|x_1 - x_2|, |y_1 - y_2|\} \quad (4-46)$$

这里采用类似 Color Correlogram 的方法，通过定义条件概率来描述图像状态矩阵中各个状态之间的空间关系特征。

设  $H_{b_i}(\mathbf{T})$  代表状态矩阵  $\mathbf{T}$  中状态为  $b_i$  ( $1 \leq i \leq L$ ) 的状态点的总个数，则状态矩阵  $\mathbf{T}$  中状态  $b_i$  出现的概率可表示为

$$h_{b_i}(\mathbf{T}) = \Pr[t \in \mathbf{T}_{b_i}] = \frac{H_{b_i}(\mathbf{T})}{n^2} \quad (4-47)$$

设  $\mathbf{D}$  表示预先定义的距离集合， $k \in \mathbf{D}$  表示状态矩阵中状态间的距离， $1 \leq i \leq L$ ， $1 \leq j \leq L$ ，从而，状态矩阵  $\mathbf{T}$  的状态相关图定义为

$$\gamma_{b_i, b_j}^{(k)}(\mathbf{T}) \triangleq \Pr[t_2 \in \mathbf{T}_{b_j}, |t_1 - t_2| = k | t_1 \in \mathbf{T}_{b_i}] \quad (4-48)$$

如果仅仅考虑图像状态矩阵中相同状态间的空间关系，此时的状态相关图称为状态自关联图 (state auto-correlogram)，其定义为

$$\beta_b^{(k)}(\mathbf{T}) = \gamma_{b,b}^{(k)}(\mathbf{T}) \quad (4-49)$$

##### 2. 状态相关图计算 (Huang J. 1998)

令

$$\Gamma_{b_i, b_j}^{(k)}(\mathbf{T}) \triangleq \left| \left\{ t_1 \in \mathbf{T}_{b_i}, t_2 \in \mathbf{T}_{b_j} \mid |t_1 - t_2| = k \right\} \right| \quad (4-50)$$

式（4-50）统计了空间距离为  $k$  且状态分别为  $b_i$ 、 $b_j$  的状态对的个数，其中  $1 \leq i \leq L, 1 \leq j \leq L$ 。

如果忽略边界条件, 设  $b^{(1)}$ 、 $b^{(2)}$  分别代表状态矩阵中任意两状态点  $t_1$ 、 $t_2$  的状态, 则

$$\Pr(b^{(1)} = b_i, b^{(2)} = b_j, |t_1 - t_2| = k) = \frac{\Gamma_{b_i, b_j}^{(k)}(\mathbf{T})}{n^2 \times 8k} \quad (4-51)$$

由于对于状态矩阵中的每一个状态点, 如果不考虑边界条件, 则和它距离为  $k$  的状态点的个数为  $8k$ , 因此在式 (4-51) 中分母表示在状态矩阵  $\mathbf{T}$  中空间距离为  $k$  的所有状态点的总个数。由此得到如下条件概率:

$$\begin{aligned} \Pr(b^{(2)} = b_j, |t_1 - t_2| = k | b^{(1)} = b_i) &= \frac{\Pr(b^{(1)} = b_i, b^{(2)} = b_j, |t_1 - t_2| = k)}{\Pr(b^{(1)} = b_i)} \\ &= \frac{n^2}{H_{b_i}(\mathbf{T})} \times \frac{\Gamma_{b_i, b_j}^{(k)}(\mathbf{T})}{n^2 \times 8k} = \frac{\Gamma_{b_i, b_j}^{(k)}(\mathbf{T})}{H_{b_i}(\mathbf{T}) \times 8k} \end{aligned} \quad (4-52)$$

式 (4-52) 即为所要计算的  $\gamma_{b_i, b_j}^{(k)}$  的值。

同应用于灰度图像纹理分析的灰度共生矩阵 (Co-Occurrence) 一样, 式 (4-52) 具有较高的计算复杂度, 为此 Jing Huang (Huang J. 1998) 给出了一种快速算法: 令  $t = (x, y)$  表示状态矩阵中的任意一状态点, 设  $\lambda_t^{b, h}(k)$  表示图像状态矩阵中在水平方向上同  $t$  间距离小于  $k$  且状态为  $b$  的状态点的数目,  $\lambda_t^{b, v}(k)$  表示图像状态矩阵中在垂直方向上同  $t$  间距离小于  $k$  且状态为  $b$  的状态点的数目, 其定义如下:

$$\lambda_t^{b, h}(k) \triangleq |\{(x+i, y) \in \mathbf{T}_b \mid 0 \leq i \leq k\}| \quad (4-53)$$

$$\lambda_t^{b, v}(k) \triangleq |\{(x, y+i) \in \mathbf{T}_b \mid 0 \leq i \leq k\}| \quad (4-54)$$

根据上述定义很容易得到

$$\lambda_{(x, y)}^{b, h}(k) = \lambda_{(x, y)}^{b, h}(k-1) + \lambda_{(x+k, y)}^{b, h}(0) \quad (4-55)$$

式中, 初始化条件  $\lambda_t^{b, h}(0) = \begin{cases} 1, t \in \mathbf{T}_b \\ 0, t \notin \mathbf{T}_b \end{cases}$ 。

根据式 (4-55), 我们很容易计算出  $\lambda_t^{b, v}$ 。

根据上述分析,  $\Gamma_{b_i, b_j}^{(k)}(\mathbf{T})$  可采用式 (4-56) 计算 (见图 4.8)

$$\begin{aligned} \Gamma_{b_i, b_j}^{(k)}(\mathbf{T}) &= \sum_{(x, y) \in \mathbf{T}_{b_i}} (\lambda_{(x-k, y+k)}^{b_j, h}(2k) + \lambda_{(x-k, y-k)}^{b_j, h}(2k) + \\ &\quad \lambda_{(x-k, y-k+1)}^{b_j, v}(2k-2) + \lambda_{(x+k, y-k+1)}^{b_j, h}(2k-2)) \end{aligned} \quad (4-56)$$

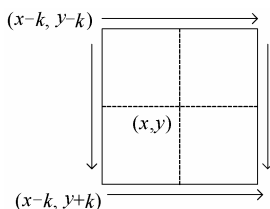


图 4.8 式 (4-56) 计算示意图

## 4.6 基于平坦度及凹凸度的描述方法

### 4.6.1 平坦度及凹凸度定义

对彩色图像来说, 图像的形状变化主要是通过图像亮度分量中像素灰度的变化来反映, 如图 4.9 所示, 其中图 (a) 为原始图像, 图 (b) ~ (d) 为其三个分量。因此, 对于某一区域子块, 如果该区域内像素间的灰度值差别较小, 那么在人眼看来, 该区域表现的就比较平坦, 如果该区域内像素间的灰度值差别较大, 则该区域表现得就比较起伏。区域内像素灰度值的这种变化可以通过分块区域内像素灰度值的标准差来反映, 因而对于某一子块, 如果其内部像素灰度值的标准差很小, 则表明该子块内像素的灰度值比较接近, 因此, 该子块对应图像中灰度变化平坦的区域, 如果其内部像素灰度值的标准差很大, 则表明该子块内像素的灰度值变化较大, 而只有包含图像边缘细节的区域子块才具有这种特性。

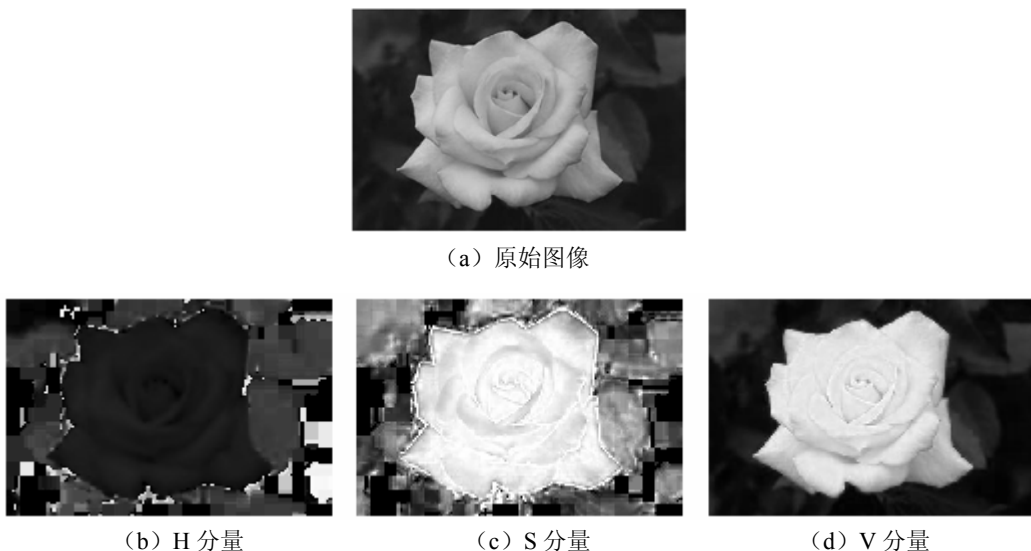


图 4.9 示例图像 HSV 分解示意图

为了描述图像形状的这种变化特征, 引入两个新的图像分块形状描述符: 平坦度和凹凸度 (孙君顶等, 2004c)。

设  $v(i, j)$  表示亮度分量中像素  $(i, j)$  的灰度值, 则有如下定义。

**定义 6:** 平坦度 图像亮度分量子块内所有像素灰度的均值定义为该子块的平坦度。定义为

$$\text{flatness} = \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m v(i, j) \quad (4-57)$$

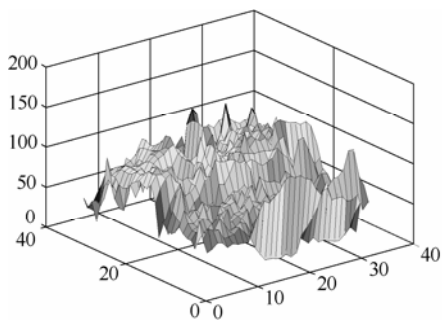
**定义 7:** 凹凸度 为了定义分块的凹凸度, 首先将亮度分量中所有像素灰度的均值  $\text{mean}$  ( $\text{mean} = \frac{1}{R_1 \times R_2} \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} v(i, j)$ ,  $R_1, R_2$  表示图像的尺寸) 看做一平面, 对于每一分块,

若  $\text{flatness} > \text{mean}$ ，则称该分块具有凸性，若  $\text{flatness} < \text{mean}$ ，则称该分块具有凹性。从而分块的凹凸度定义为

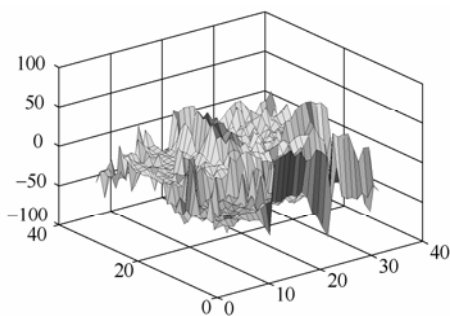
$$\text{roughness} = \begin{cases} \left[ \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m (v(i, j) - \text{flatness})^2 \right]^{\frac{1}{2}}, & \text{flatness} > \text{mean} \\ 0, & \text{flatness} = \text{mean} \\ -\left[ \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m (v(i, j) - \text{flatness})^2 \right]^{\frac{1}{2}}, & \text{flatness} < \text{mean} \end{cases} \quad (4-58)$$

式中， $l, m$  表示分块的尺寸。

图 4.10 给出了示例图 4.9 (a) 的平坦度和凹凸度的三维分布示意图，其中示例图像被划分为  $32 \times 32$  个分块。



(a) 示例图像平坦度三维分布示意图



(b) 示例图像凹凸度三维分布示意图

图 4.10 示例图像平坦度和凹凸度分布特征

### 4.6.2 形状特征量化

在提取了图像每一分块的平坦度和凹凸度特征后，接下来需要按照平坦度和凹凸度两个属性对图像的分块进行分类，这需要选用合适的量化策略对图像分块的平坦度和凹凸度进行量化。由平坦度和凹凸度的定义可以看出，平坦度反映了图像分块的基本形状，而凹凸度反映了图像分块的形状变化，因此，在对平坦度和凹凸度特征进行量化时需要采用不同的量化策略，对于平坦度采用粗量化即可，而对于反映图像形状变化的凹凸度则需要采用细量化。

为了选取合适的量化策略，我们首先对分块平坦度和凹凸度的分布特征进行了统计。在包含 10 000 幅图像的图像库中任意选取 500 幅图像，所选取的图像包括多个类型，这样可以使统计结果具有一般性，而不会局限于某一类型的图像，同时所选取的图像被划分为  $32 \times 32$  个子块。图 4.11 是对图像平坦度和凹凸度的统计结果，其中图 (a) 是平坦度统计分布特征，图 (b) 为凹凸度统计分布特征。可以看出，图像的凹凸度基本服从正态分布。在统计了图像的平坦度和凹凸度的统计分布特征后，我们就可以根据它们的分布特征，采用不同的量化手段加以量化。

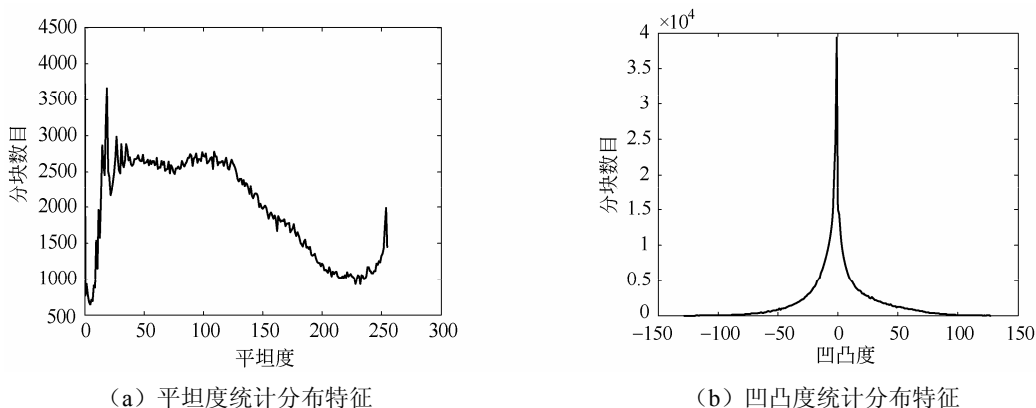


图 4.11 图像平坦度和凹凸度分布

### 1. 平坦度量化

在对平坦度进行量化时, 根据图 4.11 所示的分布特征, 提出了均匀量化和非均匀量化两种量化方法, 具体可参见参考文献(孙君顶等 2004c, 孙君顶等 2006)。

### 2. 凹凸度量

对于凹凸度, 由于其基本服从正态分布, 文中提出了采用分区间-均匀量化(孙君顶等 2004c)和等概率量化(孙君顶等 2006)两种量化方法。

#### (1) 分区间-均匀量化

由于凹凸度的分布基本服从正态分布, 设凹凸度的均值为  $\mu$ , 均方差为  $\sigma$ , 则其分布密度可表示为(盛骤等 1989)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4-59)$$

其分布函数可表示为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (4-60)$$

对于正态分布, 由于元素分布主要集中在均值  $\mu$  的两侧, 因此这里提出采用分区间-均匀的量化策略, 对于靠近均值  $\mu$  的区间进行较细的均匀量化, 对于其他区间进行较粗的均匀量化, 具体量化过程如下:

① 计算整幅图像分块凹凸度的均值  $\mu$  和均方差  $\sigma$ ;

② 由于图像凹凸度的分布基本服从正态分布, 因此凹凸度主要分布在  $[\mu - \sigma, \mu + \sigma]$  的范围内, 在对凹凸度进行量化时, 针对该区域的值进行较细的均匀量化, 该区域之外的值进行较粗的均匀量化, 量化方法为

$$Q(i, j) = \begin{cases} \frac{\text{roughness}(i, j)}{\text{step1}}, & \text{roughness}(i, j) \in [\mu - \sigma, \mu + \sigma] \\ \frac{\text{roughness}(i, j)}{\text{step2}} + \frac{(\mu + \sigma)}{\text{step1}}, & \text{roughness}(i, j) \in (\mu + \sigma, 128] \\ \frac{\text{roughness}(i, j)}{\text{step2}} + \frac{\mu - \sigma}{\text{step1}}, & \text{roughness}(i, j) \in [-128, \mu - \sigma) \end{cases} \quad (4-61)$$



式中,  $Q(i, j)$  表示对分块  $(i, j)$  凹凸度的量化结果 (实验中对  $Q(i, j)$  采用取整表示), step1、step2 表示量化步长。通过调整 step1 及 step2 的取值, 可以控制将分块划分为不同类型的精确度。为了详细的描述图像的形状变化, step1 及 step2 应取相对较小的值; 若想粗略地描述图像形状的变化, step1 及 step2 取相对较大的值即可。step1 及 step2 的取值也将会影响所提取的图像形状特征的维数变化。在量化完毕后, 根据量化结果将图像的分块划分为不同的类型, 同时采用不同类型分块的统计直方图作为图像形状特征的描述。

## (2) 等概率量化

虽然采用分区间-均匀量化也可取得较好的效果, 但是在每一划分的区间内部仍然属于刚性量化, 为了解决区间刚性量化的问题, 又提出了等概率的量化方法。等概率量化的基本思想是首先将概率区间  $[0, 1]$  进行均匀划分, 然后根据划分后的概率按逆方向计算划分的区间边界。

设  $X$  服从正态分布, 当  $\mu = 0$ ,  $\sigma = 1$  时称  $X$  服从的分布为标准正态分布, 设标准正态分布的概率密度和分布函数分别用  $\varphi(x)$  和  $\Phi(x)$  表示, 即

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (4-62)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (4-63)$$

这里引进标准正态分布是因为标准正态分布函数  $\Phi(x)$  的取值可以通过查找相应的标准正态分布函数表获取, 从而避免复杂的运算, 而且对于正态分布也可通过下面的定理将其转换为标准正态分布 (盛骤等 1989)。

**定理 2:** 若  $X \sim N(\mu, \sigma^2)$ , 则  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ , 即  $Z$  服从标准正态分布。

于是, 若  $X \sim N(\mu, \sigma^2)$ , 则它的分布函数  $F(x)$  为

$$F(x) = P\{X \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (4-64)$$

从而, 对于任意的区间  $[x_1, x_2]$ , 有

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\left\{\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right\} \\ &= \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right) \end{aligned} \quad (4-65)$$

因此, 在对图像的凹凸度进行量化时, 首先将其概率区间  $[0, 1]$  均匀划分为  $n_r$  个区间, 然后根据其概率分布求每一分界点所对应的值。具体过程如下:

① 首先将凹凸度的分布由正态分布转换为标准正态分布;

② 由于正态分布关于  $x = \mu$  对称, 因此, 只需将概率区间  $[0.5, 1]$  划分为  $\frac{n_r}{2}$  个区间即可, 从而概率区间被划分为  $\left\{0.5, 0.5 + \frac{1}{n_r}, \dots, x_i, \dots, 1 - \frac{1}{n_r}\right\}$ ;

③ 根据式 (4-65) 及所划分的概率区间计算相应的分界点的值:

$$\Phi\left(\frac{x_1 - \mu}{\sigma}\right) = 0.5 \Leftrightarrow x_1 = \mu + \sigma\Phi^{-1}(0.5) = \mu \quad (4-66)$$

$$\Phi\left(\frac{x_2 - \mu}{\sigma}\right) = 0.5 + \frac{1}{n_r} \Leftrightarrow x_2 = \mu + \sigma\Phi^{-1}\left(0.5 + \frac{1}{n_r}\right) \quad (4-67)$$

$$\Phi\left(\frac{x_i - \mu}{\sigma}\right) = 0.5 + \frac{i}{n_r} \Leftrightarrow x_i = \mu + \sigma\Phi^{-1}\left(0.5 + \frac{i}{n_r}\right) \quad (4-68)$$

$$\Phi\left(\frac{\frac{x_{n_r-1}}{2} - \mu}{\sigma}\right) = 1 - \frac{1}{n_r} \Leftrightarrow \frac{x_{n_r-1}}{2} = \mu + \sigma\Phi^{-1}\left(1 - \frac{1}{n_r}\right) \quad (4-69)$$

而  $\Phi^{-1}$  的取值可以通过查找标准正态分布表取得, 从而降低了计算复杂度。

④ 根据正态分布的对称特性, 以及上述计算得到的分界点, 即可得到在概率区间[0, 0.5]的分界点。

量化完成后, 设平坦度被量化为  $n_f$  个级别, 从而图像的分块被划分为  $n_f \times n_r$  个类别, 按照以上的量化级数, 我们把平坦度 (flatness) 和凹凸度 (Roughness) 两个分量合并为一个一维的特征向量

$$h_s = n_r \times \text{flatness}' + \text{roughness}' \quad (4-70)$$

式中, Flatness' 和 Roughness' 为平坦度和凹凸度的量化结果。这样平坦度 (Flatness) 和凹凸度 (Roughness) 两个分量在一维向量上分布开来, 根据式 (4-70),  $h_s$  的取值范围为[0,  $n_f \times n_r - 1$ ]。也就是说, 根据量化后的结果可以统计得到  $n_f \times n_r$  柄的一维直方图, 该直方图即可用来描述形状特征。

## 4.7 基于信息熵的描述方法

对于数字图像而言, 图像由像素组成, 不同灰度的像素出现次数的不同及其分布空间位置的不同, 使得图像呈现不同的形状。因此, 不同形状的图像所包含的熵也是不尽相同的, 故可以用熵描述图像的形状特征, 这里称为 EBIR (Entropy-Based Image Retrieval) 算法 (孙君顶等 2004b, 孙君顶等 2007)。

### 4.7.1 图像信息熵定义

灰度直方图是灰度级的函数, 它表示图像中具有某种灰度级的像素的个数, 反映图像中每种灰度出现的频率。设图像像素灰度级为集合  $\mathbf{M}$ , 对于灰度图像  $\mathbf{I}$ , 其灰度直方图可定义为

$$p_m = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(I(i, j) - m) \quad (4-71)$$

式中,  $\delta(I(i, j) - m) = \begin{cases} 1, & I(i, j) = m \\ 0, & I(i, j) \neq m \end{cases}, \forall m \in \mathbf{M}$ ,  $M \times N$  表示图像的尺寸。

从概率的观点来理解, 灰度出现的频率可看做灰度出现的概率, 这样直方图就对应于灰度图像的概率密度函数 (Probability Density Function)。式 (4-71) 对不同灰度出现的概率密度进行了归一化处理, 从而有  $0 \leq p_m \leq 1$ , 且满足  $\sum_{m \in \mathbf{M}} p_m = 1$ , 因此向量  $(p_1, p_2, \dots, p_m, \dots)$

可称作图像灰度分布的概率向量, 根据信息熵的定义, 图像的信息熵定义为

$$e = - \sum_{m \in M} p_m \log_2(p_m) \quad (4-72)$$

由熵的定义和性质可以知道，当一幅图像中只包含一种灰度级时，图像的信息熵为零；相反，当图像中所有灰度级出现的概率相等时，图像的信息熵最大。

具体到某一幅图像，其所包含的灰度级  $M'$  往往为整个灰度集合  $M$  的子集，即  $M' \subseteq M$ 。同时，某些灰度在一幅图像中出现的概率往往是很小的，根据熵的可扩展特性，这部分灰度对图像的整体信息熵的影响是很小的，因此在计算图像信息熵时，这部分灰度的影响可忽略不计。图 4.12 给出了六幅示例图像，它们的信息熵如表 4.1 所示。如果仅仅考虑 51% 的灰度级，针对图 4.12 所示的六幅图像，其信息熵平均保持为原来的 90% 以上，具体取值如表 4.2 所示。

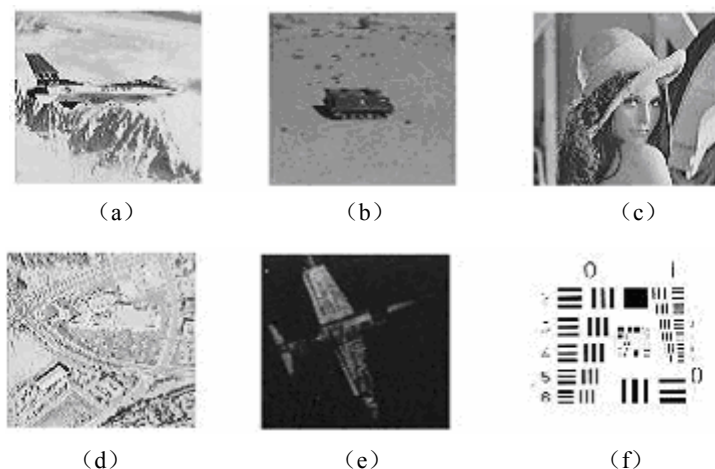


图 4.12 示例图像

表 4.1 示例图像的信息熵

图像	(a)	(b)	(c)	(d)	(e)	(f)
熵	6.7247	5.8735	7.4446	7.0089	5.5695	2.5725

表 4.2 根据熵的扩展特性处理后示例图像的信息熵

图像	(a)	(b)	(c)	(d)	(e)	(f)
熵	6.4382	5.6651	7.3737	6.7084	5.2156	1.8598

### 4.7.2 图像的单元熵

通过式 (4-72) 对图像信息熵的定义可知，图像的信息熵表现了图像灰度分布的全局统计特性。图像信息熵的大小只与图像中各灰度出现的概率有关，具有相同概率分布的两幅图像具有相同的信息熵。同时，图像的信息熵与图像的全局直方图一样仅仅考虑了图像的全局统计信息，丢弃了图像的空间分布信息，因此，具有相同信息熵的图像可能在视觉上是完全不同的，所以图像的全局信息熵不足以反映出图像间的差异。

我们知道，不同图像之所以能呈现出不同的形状，一方面是因为不同灰度值的像素出现的频数不同；另一方面也跟不同灰度的像素的空间分布特征有关。为了提取图像的形状

特征, 在文献 (Lu G.J. et al 1999, Safar M. et al 2000, Chakrabarti K. et al 2000) 中, 引入了网格描述符 GD (Grid Descriptor), 如图 4.13 (Chakrabarti K. et al 2000) 所示。在 GD 中, 图像首先被投射到具有固定分辨率的网格上, 若网格单元被图像内容覆盖 (或被覆盖的程度超过一定阈值), 则该网格赋值为 “1”, 否则赋值为 “0”, 然后通过对此网格进行从左到右、从上到下的扫描, 形成一串二进制序列, 并通过该二进制序列来描述图像的形状并进行图像间的相似性检索。

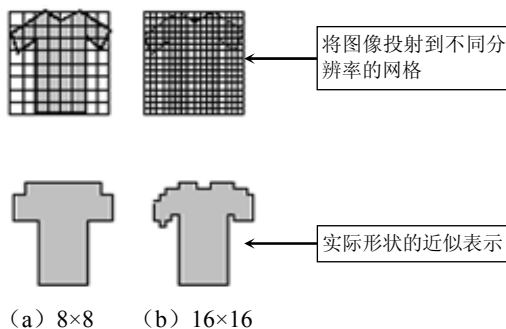


图 4.13 示例图像固定分辨率表示

但是在上述基于网格的描述方法中, 完全抛开了图像的灰度信息, 仅仅采用简单的赋 “0”、赋 “1” 来描述每一单元格的特征。基于此, 这里所提出的 EBIR 算法在基于 GD 方法的基础上, 采用图像局部灰度的统计特性来描述图像的形状信息。同 GD 方法一样, 首先将图像投射到一固定分辨率的网格上 (网格的分辨率随图像尺寸变化而变化), 在对每一网格单元进行赋值时不是采用 GD 方法, 而是将包含该单元格灰度信息的信息熵作为该单元格的特征描述。为了描述图像单元格的这种特征, 文中引入了单元熵的概念, 设  $e_{ij}$  表示在图像空间中坐标为  $(i, j)$  网格单元的单元熵, 则  $e_{ij}$  定义如下:

$$e_{ij} = - \sum_{m \in M} p'_m \log(p'_m) \quad (4-73)$$

式中,  $p'_m$  表示单元格内灰度为  $m$  的像素出现的概率。

采用单元熵对图像单元格的特征进行表示后, 整幅图像就变成了一个由不同单元熵构成的熵矩阵, 该熵矩阵包含了图像的全局和局部形状特性。采用熵矩阵来描述图像特征具有如下特点:

(1) 从全局图像来说, 由于单元熵代表了图像局部单元的统计特征, 因此由所有单元熵构成的熵矩阵不仅描述了图像的全局特性也反映了图像的空间分布特性, 这解决了采用图像全局熵描述图像特征所造成的熵相同而形状不同的问题, 也解决了由于熵的对称性所造成的影响。

(2) 从局部上讲, 由于各单元熵体现了图像局部灰度的统计特性, 因此 EBIR 算法充分利用了图像的灰度信息来描述图像的特征, 加强了图像检索的准确性。

(3) 利用熵矩阵描述形状, 其准确度与所采用的单元格的分辨率有关, 理论上分辨率越高对图像形状描述越准确, 但在这种情况下, 图像特征熵矩阵的维数会增大, 存储图像特征所需的空间也会增大, 同时还会造成图像检索速度的下降。实验表明对  $256 \times 256$  的图像采用  $16 \times 16$  的分辨率即可取得较好的效果。

虽然图像的熵矩阵可以用来区分不同形状的图像，但如果直接以熵矩阵作为图像的形状特征用于图像的相似性检索，还存在下列问题：

(1) 维数高。很明显，所采用网格的分辨率与熵矩阵的维数密切相关。为了保证能尽量准确的体现原图像的形状特征，网格的分辨率不易太小，因此熵矩阵的维数将会很大，从而为图像的相似性度量带来困难。

(2) 在许多情况下，我们希望提取出的图像形状特征在平移、旋转、尺度等条件的变化下是一个不变量，显然，熵矩阵并不满足这种要求。

为此，在利用计算得到图像的熵矩阵后，我们采用不同的方法对熵矩阵进行处理，以克服上面提到的问题。

### 4.7.3 利用熵矩阵的特征值向量进行检索

图像的特征采用熵矩阵描述后，即可采用熵矩阵来进行图像间的相似性度量，但采用熵矩阵作为图像的特征描述存在特征维数高的问题。图像中相邻的单元格之间具有一定的相关性，因此，在熵矩阵中相邻的单元熵之间也具有一定的信息冗余。为了消除这种信息冗余，我们采用了熵矩阵的特征值向量来代替熵矩阵作为图像的索引特征，同时为了降低所提取的特征值向量的维数，我们采用特征值向量中模最大的几个特征值来进行图像间的相似性度量。对于特征值向量中模较小的特征值，由于其包含的信息量很少，忽略掉这些特征值不会对检索的结果造成太大的影响，而同时由于采用较少的特征值参与图像间的相似性计算，从而加快了图像检索的速度（孙君顶等 2004b）。

利用熵矩阵特征值向量进行检索的主要过程如下。

(1) 将图像投射到分辨率为  $n \times n$  的网格上，针对每一单元格，计算不同灰度级的概率分布向量  $(p_0, p_1, \dots, p_{255})$ ，然后采用式 (4-73) 计算单元格的单元熵  $e_{ij}$ ，从而求取图像的熵矩阵  $E$ 。根据上述熵的可扩展特性可知，如果图像单元格内的某一灰度的概率分布很小，则该灰度对单元格的信息熵的影响可忽略不计。因此，在计算图像的单元熵时可预先设定一阈值  $\alpha$ ，若  $p_m < \alpha$ ，图像的单元熵在计算时该灰度级可忽略不计。

$$E = \{e_{ij}\} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix} \quad (1 \leq i \leq n, 1 \leq j \leq n) \quad (4-74)$$

(2) 将熵矩阵  $E$  的特征值按照模的大小排列成一维向量  $(\lambda_1, \lambda_2, \dots, \lambda_n)$ （其中， $\|\lambda_1\| \geq \|\lambda_2\| \geq \dots \geq \|\lambda_n\|$ ），提取特征值向量的前  $t$  个主特征值作为图像的熵描述符。 $t$  取值满足下述条件：

$$\sum_{l=1}^t \lambda_l / \sum_{l=1}^n \lambda_l \geq \beta \quad (4-75)$$

式中， $\beta$  为预定义阈值，且  $\beta \in [0, 1]$ 。

(3) 对于提取的特征值向量，利用欧氏距离对图像间的相似度进行度量，即

$$d = \sqrt{\sum_{l=1}^t \|\lambda_l - \lambda_l^{(i)}\|^2} \quad (4-76)$$

式中， $\lambda_l$  表示待检索图像的第  $l$  个特征值， $\lambda_l^{(i)}$  表示用于检索的第  $i$  幅图像的第  $l$  个特征

值。计算出图像间的距离后，利用两幅图像间的距离判断出两幅图像的相似程度，并依照事先设定的阈值检索出与例子图像相似的图像来。

#### 4.7.4 利用熵矩阵的不变矩进行检索

不变矩是一个重要的基于区域的形状特征。不变矩是通过所有属于区域内的像素点计算出来的，因而不受噪声等因素的影响（章毓晋 1999）；并且由于不变矩特征具有良好的尺度、平移和旋转不变性，因而在基于形状的图像检索中取得了非常好的检索性能（黄元元 2003，姚玉荣等 2000）。这里我们在熵矩阵的基础上，采用不变矩来描述图像特征（孙君顶等 2007），有关不变矩的详细介绍可参考 4.3.1 节，这里不再详细介绍。

由上述方法得到的不变矩特征向量，每一个元素的物理意义都是不同的，因此他们的幅度也存在较大的差异。如果直接用它们进行图像间的相似性度量，会产生很大的偏差，故必须通过归一化处理来消除这种偏差（洪安祥 2003）。

设图像的特征向量为  $\mathbf{R} = \{r_1, r_2, \dots, r_K\}$ ， $K$  表示特征的维数，特征内部归一化过程的目的在于使特征向量  $\mathbf{R}$  的各分量  $r_k (1 \leq k \leq K)$  具有同等的重要性。同第 3 章所述的图像特征间的归一化一样，我们不可能对图像库中的所有图像进行统计，只是采用训练集的方式统计特征向量中每一维的分布特征。假设我们从数据库中选取  $M$  幅图像作为训练集，设  $i$  为训练集中图像的索引值，那么我们采用  $\mathbf{R}_i$  表示第  $i$  幅图像的特征向量，即

$$\mathbf{R}_i = (r_{i,1}, r_{i,2}, \dots, r_{i,k}, \dots, r_{i,K}) \quad (4-77)$$

如果我们将所有的  $\mathbf{R}_i (1 \leq i \leq M)$  叠加在一起，就可以得到维数为  $M \times K$  的矩阵，即

$$\mathbf{V} = [v_{i,k}], 1 \leq i \leq M, 1 \leq k \leq K \quad (4-78)$$

式中， $v_{i,k}$  是特征向量  $\mathbf{R}_i$ （对应于第  $i$  幅图像）的第  $k$  个分量。矩阵的第  $k$  列是维数为  $M$  的一个列向量，记为  $\mathbf{V}_k$ 。我们的目的是将每列中的元素都归一化到一致的值域范围中，以保证在计算两个向量的相似度时，各分量具有同等重要性。

假设列向量  $\mathbf{V}_k$  是一个高斯数列，我们首先计算该数列的均值  $\mu_k$  和标准差  $\sigma_k$ ，然后采用下式来进行归一化操作

$$v_{i,k} \leftarrow \frac{v_{i,k} - \mu_k}{3\sigma_k} \quad (4-79)$$

根据  $3-\sigma$  规则（数列中各值落入区间  $[-1, 1]$  范围内的概率约为 99%），在实际应用中我们可以认为数列中的所有元素已落入区间  $[-1, 1]$  范围中了，对于非此范围中的数值可以简单地对应到  $-1$  和  $1$  上。

#### 4.7.5 算法特性

实验对上述算法的尺度不变性、旋转不变性、平移不变性、图像灰度变化对 EBIR 算法的影响以及噪声存在对 EBIR 算法的影响进行了验证（孙君顶等 2004，孙君顶 2005，孙君顶等 2007）。

针对式（4-75）中预先定义的阈值  $\beta (\beta \in [0, 1])$ ，若  $\beta = 1$ ，表明在检索时取熵矩阵的所有特征值组成检索向量来进行图像检索；若  $\beta < 1$ ，表明在检索时从特征值向量中取部分主特征值组成检索向量来进行图像检索。由于所提取的图像特征主要集中在特征向量中靠前的几个特征值上，随后的特征值所包含的信息很少，对不同类别的图像进行统计发

现, 在特征值向量中靠前的 5~8 个特征值所包含的信息占整个特征值向量所包含信息的 90% 以上, 采用这 5~8 个特征值组成检索向量即可进行有效的图像检索。在下列实验中, 取  $\beta = 0.95$ , 网格的分辨率取  $16 \times 16$ 。

### 1. 尺度不变性

为了简化计算, 这里采用二值图像来验证 EBIR 算法的尺度不变性。如图 4.14 所示, 其中图 (a)、(b) 和 (c) 为不同尺度的相同图像, 图 (d) 为不同于图 (a)、(b) 和 (c) 的另外一幅图像。采用熵矩阵特征值向量进行距离度量, 4 幅图像间的相互距离如表 4.3 所示。采用熵矩阵不变矩向量进行距离度量, 4 幅图像间的相互距离如表 4.4 所示。从表 4.3 和表 4.4 可以看出, 不同尺度的相同图像之间的距离远小于它们与视觉不同图像之间的距离, 因此 EBIR 算法具有尺度不变性, 并且采用熵矩阵特征值向量的效果相对好一些。

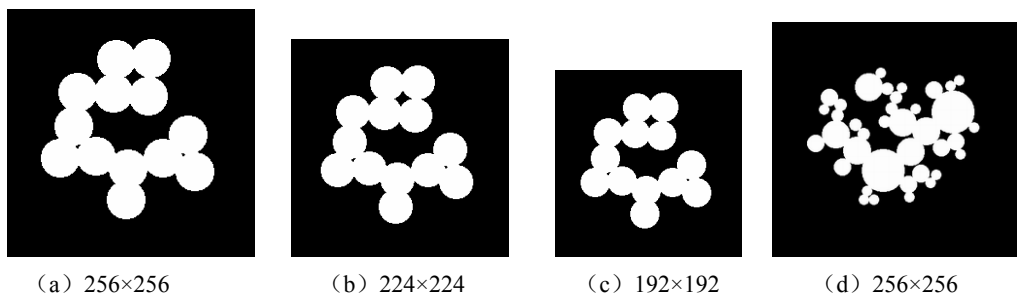


图 4.14 不同尺度二值图像

表 4.3 采用熵矩阵特征值进行距离度量结果

	(a)	(b)	(c)	(d)
(a)	0	0.0515	0.0653	1.5351
(b)	0.0515	0	0.0319	1.498
(c)	0.0653	0.0319	0	1.4871
(d)	1.5351	1.498	1.4871	0

表 4.4 采用熵矩阵不变矩进行距离度量结果

	(a)	(b)	(c)	(d)
(a)	0	0.248	0.3267	1.1471
(b)	0.248	0	0.1615	1.1335
(c)	0.3267	0.1615	0	1.1031
(d)	1.1471	1.1335	1.1031	0

### 2. 旋转不变性

如图 4.15 所示, 图 (a) 为原始图像, 图 (b) ~ (h) 分别为图 (a) 在水平、垂直方向按  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$  旋转的结果, 图 (i) 为不同于图 (a) 的另外一幅图像。采用熵矩阵的特征值向量进行检索, 这 9 幅图像间的距离如表 4.5 所示, 采用熵矩阵的不变矩特征进行检索, 这 9 幅图像间的距离如表 4.6 所示。从表 4.5 和表 4.6 可以看出, 旋转图像

之间的距离远小于它们与视觉不同图像之间的距离。因此,对于此类旋转图像,无论是采用熵矩阵的特征值向量还是采用熵矩阵的不变矩向量进行检索,均可取得较好的效果。

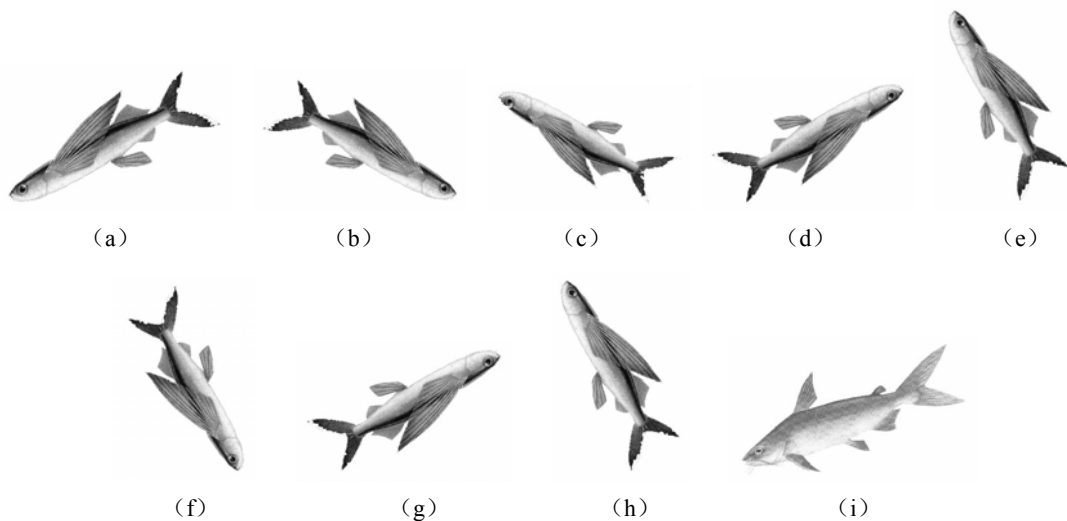


图 4.15 示例图像及其旋转图像

表 4.5 采用熵矩阵特征值进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
(a)	0	1.4328	1.6306	0.9544	1.39	1.4338	0.9544	1.39	10.887
(b)	1.4328	0	1.3062	1.8588	0.8873	0.2212	1.8588	0.8873	11.065
(c)	1.6306	1.3062	0	2.0011	1.2686	1.1864	2.0011	1.2686	10.462
(d)	0.9544	1.8588	2.0011	0	1.749	1.8686	0	1.749	10.73
(e)	1.39	0.8873	1.2686	1.749	0	0.8883	1.749	0	11.132
(f)	1.4338	0.2212	1.1864	1.8686	0.8883	0	1.8686	0.8883	11.068
(g)	0.9544	1.8588	2.0011	0	1.749	1.8686	0	1.749	10.73
(h)	1.39	0.8873	1.2686	1.749	0	0.8883	1.749	0	11.132
(i)	10.887	11.065	10.462	10.73	11.132	11.068	10.73	11.132	0

表 4.6 采用熵矩阵不变矩进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
(a)	0	0.041	0.1981	0.1019	0.1717	0.0714	0.1019	0.1717	2.5511
(b)	0.041	0	0.182	0.0727	0.1391	0.0447	0.0727	0.1391	2.6694
(c)	0.1981	0.182	0	0.188	0.164	0.1561	0.188	0.164	2.555
(d)	0.1019	0.0727	0.188	0	0.0941	0.091	0	0.0941	2.6969
(e)	0.1717	0.1391	0.164	0.0941	0	0.1234	0.0941	0	2.6637
(f)	0.0714	0.0447	0.1561	0.091	0.1234	0	0.091	0.1234	2.6434
(g)	0.1019	0.0727	0.188	0	0.0941	0.091	0	0.0941	2.6969
(h)	0.1717	0.1391	0.164	0.0941	0	0.1234	0.0941	0	2.6637
(i)	2.5511	2.6694	2.555	2.6969	2.6637	2.6434	2.6969	2.6637	0



图 4.16 所示的 8 幅图像为图像 (a) 按照  $15^\circ$ 、 $30^\circ$ 、 $45^\circ$ 、 $60^\circ$ 、 $135^\circ$ 、 $150^\circ$  和  $175^\circ$  进行旋转所得到的图像。采用熵矩阵的特征值向量, 旋转图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.7 所示。采用熵矩阵的不变矩特征, 旋转图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.8 所示。从表 4.7 和表 4.8 可以看出, 相对于表 4.5 和表 4.6 来说, 图像间的距离变大了, 也就是说, EBIR 算法针对此类旋转图像的检索效果没有对图 4.15 所示旋转特性图像的检索效果好, 但总体来说, 针对此类旋转的图像也能有效的检索。同时可以看出, 针对此类旋转图像, 采用熵矩阵不变矩特征进行检索的效果明显优于采用熵矩阵特征值向量进行检索的效果。

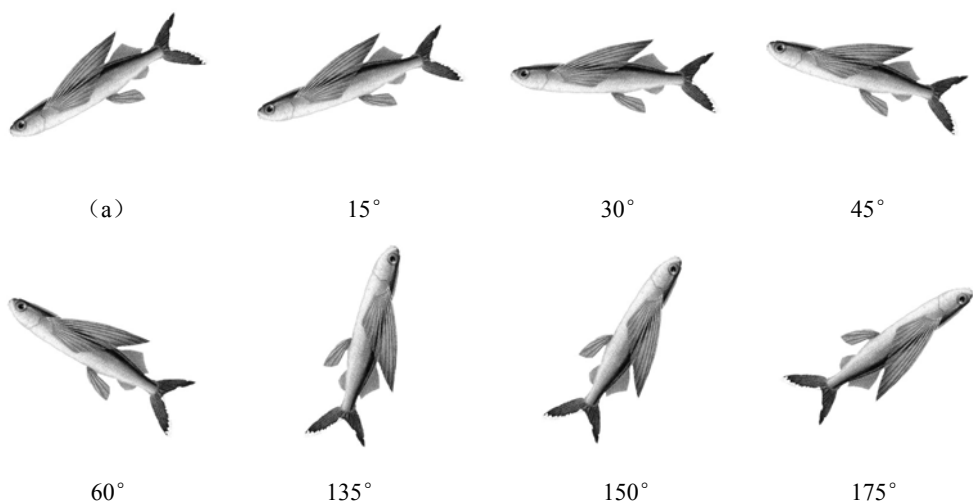


图 4.16 示例图像及其旋转图像

表 4.7 采用熵矩阵特征值进行距离度量结果

	(a)	$15^\circ$	$30^\circ$	$45^\circ$	$60^\circ$	$135^\circ$	$150^\circ$	$175^\circ$	(i)
(a)	0	3.5865	4.2198	3.7184	5.8363	4.5075	4.8882	6.803	10.887
$15^\circ$	3.5865	0	2.5388	1.5605	7.2713	3.0976	6.2721	8.5519	9.6478
$30^\circ$	4.2198	2.5388	0	1.4787	7.1781	1.3802	7.0901	9.2801	10.462
$45^\circ$	3.7184	1.5605	1.4787	0	7.6283	2.3743	6.9091	9.2701	9.1496
$60^\circ$	5.8363	7.2713	7.1781	7.6283	0	6.8595	4.7116	4.0356	11.89
$135^\circ$	4.5075	3.0976	1.3802	2.3743	6.8595	0	7.5857	9.344	9.0834
$150^\circ$	4.8882	6.2721	7.0901	6.9091	4.7116	7.5857	0	3.7614	10.73
$175^\circ$	6.803	8.5519	9.2801	9.2701	4.0356	9.344	3.7614	0	12.786
(i)	10.887	9.6478	8.7616	9.1496	11.89	9.0834	11.709	12.786	0

表 4.8 采用熵矩阵不变矩进行距离度量结果

	(a)	$15^\circ$	$30^\circ$	$45^\circ$	$60^\circ$	$135^\circ$	$150^\circ$	$175^\circ$	(i)
(a)	0	0.288	0.2425	0.3563	0.1424	0.351	0.09	0.2942	2.5511
$15^\circ$	0.288	0	0.4173	0.4211	0.2709	0.3629	0.28	0.2359	2.4568
$30^\circ$	0.2425	0.4173	0	0.234	0.1845	0.3106	0.1898	0.4982	2.5613

(续表)

	(a)	15°	30°	45°	60°	135°	150°	175°	(i)
45°	0.3563	0.4211	0.234	0	0.3248	0.1427	0.3231	0.5316	2.6089
60°	0.1424	0.2709	0.1845	0.3248	0	0.3464	0.0656	0.3612	2.4904
135°	0.351	0.3629	0.3106	0.1427	0.3464	0	0.3349	0.4696	2.6264
150°	0.09	0.28	0.1898	0.3231	0.0656	0.3349	0	0.344	2.5281
175°	0.2942	0.2359	0.4982	0.5316	0.3612	0.4696	0.344	0	2.4429
(i)	2.5511	2.4568	2.5613	2.6089	2.4904	2.6264	2.5281	2.4429	0

### 3. 平移不变性

如图 4.17 所示的 (b)、(c)、(d) 和 (e) 4 幅图像为图 4.15 (a) 的不同平移图像。采用熵矩阵的特征值向量进行检索, 图 4.15 (a) 及平移图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.9 所示。采用熵矩阵的不变矩特征, 图 4.15 (a) 及平移图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.10 所示。由表 4.9 和表 4.10 可以看出, 图像的平移对采用熵矩阵的特征值向量进行检索的影响较大, 采用熵矩阵不变矩特征进行检索的效果要优于采用熵矩阵特征值向量进行检索的效果。



图 4.17 图 4.15 (a) 的 4 幅平移图像

表 4.9 采用熵矩阵特征值进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(i)
(a)	0	6.8918	3.9443	3.3298	7.2237	10.887
(b)	6.8918	0	6.1971	5.5137	1.9228	8.2189
(c)	3.9443	6.1971	0	1.7244	6.5515	7.7513
(d)	3.3298	5.5137	1.7244	0	5.9252	7.7521
(e)	7.2237	1.9228	6.5515	5.9252	0	9.4417
(i)	10.887	8.2189	7.7513	7.7521	9.4417	0

表 4.10 采用熵矩阵不变矩进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(i)
(a)	0	0.8175	0.3856	0.5112	0.8132	2.5511
(b)	0.8175	0	0.4416	0.4526	0.348	1.6119
(c)	0.3856	0.4416	0	0.3118	0.5296	1.4317
(d)	0.5112	0.4526	0.3118	0	0.3769	1.6812
(e)	0.8132	0.348	0.5296	0.3769	0	1.8273
(i)	2.5511	1.6119	1.4317	1.6812	1.8273	0

#### 4. 图像灰度变化对算法的影响

如图 4.18 所示的 (b)、(c)、(d) 和 (e) 4 幅图像分别为图 4.15 (a) 的不同灰度变换后的图像。采用熵矩阵的特征值进行检索, 图 4.15 (a) 及不同灰度图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.11 所示。采用熵矩阵的不变矩特征, 图 4.15 (a) 及不同灰度图像与视觉特征不同的另一幅图像 (图 4.15 (i)) 间的距离如表 4.12 所示。从表 4.11 和表 4.12 可以看出, EBIR 算法对于灰度发生变化的图像也可有效的进行检索, 并且采用熵矩阵特征值向量进行检索的效果优于采用熵矩阵不变矩特征进行检索的效果。



图 4.18 图 4.15 (a) 的不同灰度的图像

表 4.11 采用熵矩阵特征值进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(i)
(a)	0	1.7501	0.8872	2.875	2.2874	10.887
(b)	1.7501	0	1.4161	1.4769	1.0202	9.3787
(c)	0.8872	1.4161	0	2.344	1.81	10.137
(d)	2.875	1.4769	2.344	0	0.70311	8.8179
(e)	2.2874	1.0202	1.81	0.7031	0	9.1457
(i)	10.887	9.3787	10.137	8.8179	9.1457	0

表 4.12 采用熵矩阵不变矩进行距离度量结果

	(a)	(b)	(c)	(d)	(e)	(i)
(a)	0	0.4993	0.2889	0.6667	0.9703	2.5511
(b)	0.4993	0	0.2329	0.3277	0.8044	1.5541
(c)	0.2889	0.2329	0	0.47	0.8854	1.4469
(d)	0.6667	0.3277	0.47	0	0.6344	1.5781
(e)	0.9703	0.8044	0.8854	0.6344	0	1.6154
(i)	2.5511	1.5541	1.4469	1.5781	1.6154	0

#### 5. 图像噪声对算法的影响

如图 4.19 所示, 图 (a) 表示原始图像, 图 (b)、(c) 和 (d) 分别表示图 (a) 采用“gaussian”噪声、“speckle”噪声及“salt & pepper”噪声加噪后的图像, 图 (e) 为不同于图 (a)、(b)、(c) 和 (d) 的另外一幅图像。采用熵矩阵的特征值向量进行检索, 这 5 幅图像间的距离如表 4.13 所示, 采用熵矩阵的不变矩特征进行检索, 这 5 幅图像间的距离如表 4.15 所示。由表 4.13 中计算出的距离可以看出, 采用熵矩阵特征值向量进行检索的方法对噪声反映比较敏感。根据熵的可扩展特性, 如果仅仅考虑图像中概率密度大于 0.8% 的灰度进行计算, 则 5 幅图像间的距离如表 4.14 所示。由该表所计算的图像间的距离可以看出, 采用上述处理方法后, 图 (a) 与噪声图 (b)、(c) 和 (d) 间的距离明显小于它们同图 (e) 间的距离, 因此采用熵的扩展特性可有效的消除噪声对检索效果带来的

影响。经过多次实验发现, 4.6.3 节中阈值  $\alpha$  的选取对抑制图像噪声的影响很重要, 如果  $\alpha$  选取不合适, 对噪声的抑制将会大大减弱。而图像噪声对不变矩特征影响较小, 我们也可以从表 4.15 的距离度量结果中发现这一点。因此, 对于噪声图像, 采用熵矩阵的不变矩特征检索效果相对好一些。

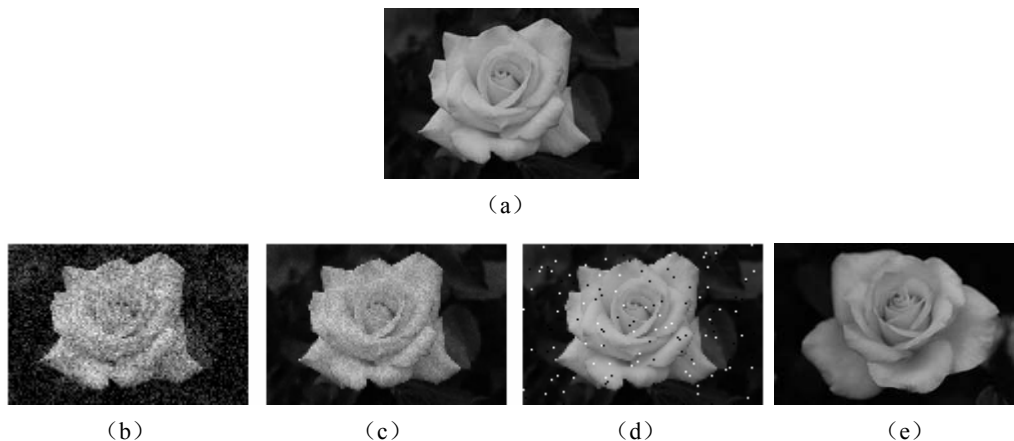


图 4.19 示例图像及加噪声后图像

表 4.13 采用熵矩阵特征值进行距离度量结果

	(a)	(b)	(c)	(d)	(e)
(a)	0	24.319	5.2325	8.7929	6.7922
(b)	24.319	0	19.329	15.696	30.756
(c)	5.2325	19.329	0	4.0433	11.575
(d)	8.7929	15.696	4.0433	0	15.252
(e)	6.7922	30.756	11.575	15.252	0

表 4.14 采用熵的扩展特性后图像间的距离

	(a)	(b)	(c)	(d)	(e)
(a)	0	2.1245	2.0366	3.0352	6.5773
(b)	2.1245	0	2.5609	2.6108	6.9384
(c)	2.0366	2.5609	0	1.7165	8.2943
(d)	3.0352	2.6108	1.7165	0	9.2125
(e)	6.5773	6.9384	8.2943	9.2125	0

表 4.15 采用熵矩阵不变矩进行距离度量结果

	(a)	(b)	(c)	(d)	(e)
(a)	0	0.4209	0.4868	0.5194	1.5401
(b)	0.4209	0	0.6239	0.589	1.3268
(c)	0.4868	0.6239	0	0.16729	1.4617
(d)	0.5194	0.589	0.16729	0	1.392
(e)	1.5401	1.3268	1.4617	1.392	0

## 4.8 基于方向链码的描述方法

### 4.8.1 基于链码的形状检索

在上文 4.2.1 节已经提到,链码是一种有效的轮廓描述方法,广泛的应用到形状的识别和编码技术中。但有关链码技术在图像检索中的应用相对还较少,常用的方法有链码直方图(Iivarinen J. et al 1996)及最小和统计方向码(王小玲等 2006)。本节就来讨论链码技术在图像检索中的应用。

#### 1. 链码直方图(CCH)

针对方向链码,起点的变化、尺度的变化、目标的旋转等都会引起链码串的变化,而在对形状进行描述时,关键就是要使所选择的描述符满足尺度、旋转和平移不变性。为了解决该问题,文献(Iivarinen J. et al 1996)提出采用链码直方图来描述链码特征并用于图像检索,链码直方图的定义为

$$h_i = \frac{n_i}{N} \quad (4-80)$$

式中,  $n_i$  表示链码串中  $i$  方向链码的个数,  $N$  表示链码串中所有链码的数目。可以看出,链码直方图反映了不同方向链码在链码串中出现的概率,是一种统计特征,与起点的选择无关,并且具有尺度及平移不变性,但不具备旋转不变性。

图 4.20 给出了不同形状的链码直方图示例。图 4.21 给出了 26 个英文字母及其链码直方图。可以看出,链码直方图在一定程度上描述了图像的轮廓特征。

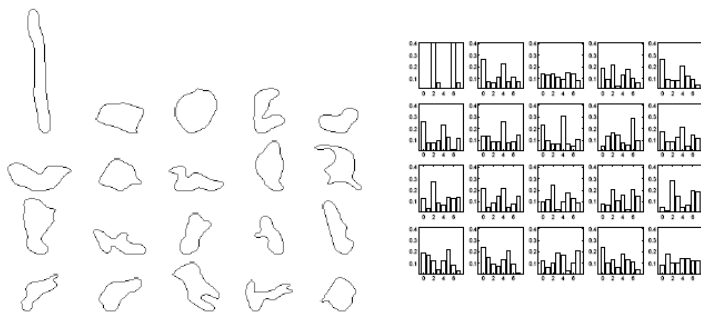


图 4.20 不同形状及其链码直方图 (Iivarinen J. et al 1996)

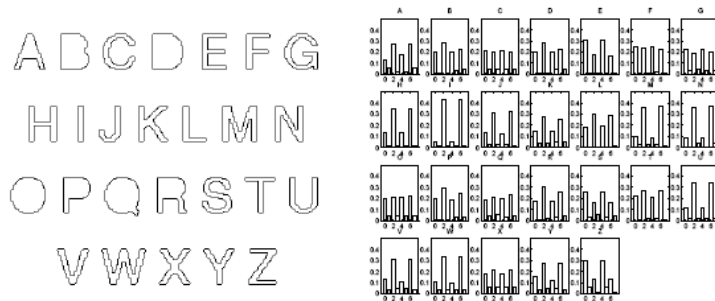


图 4.21 26 个英文字母及其链码直方图 (Iivarinen J. et al 1996)

## 2. 最小和统计方向码 (MSSDC)

首先给出统计方向码的定义。沿顺时针方向,对象轮廓方向码定义为

$$\mathbf{X} = (x_0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7) \quad (4-81)$$

式中,  $x_i$  是边缘上具有方向码  $i$  的像素数量, 故统计方向码与起始点无关, 并且具备了天然的对图像平移的不变性。构造方向代码向量, 表示链码的 8 各方向, 即

$$\mathbf{D}=[0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7]^T \quad (4-82)$$

设对象轮廓的初始方向码表示为  $\mathbf{X}_0 = (x_0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7)$  (顺时针方向), 当图像顺时针依次旋转  $45^\circ$ , 统计方向码的变化有如下规律:

$$\mathbf{X}_i = \{x_{(i+0) \oplus 8} \quad x_{(i+1) \oplus 8} \quad x_{(i+2) \oplus 8} \quad x_{(i+3) \oplus 8} \quad x_{(i+4) \oplus 8} \quad x_{(i+5) \oplus 8} \quad x_{(i+6) \oplus 8} \quad x_{(i+7) \oplus 8} \mid i=0,1,\dots,7\} \quad (4-83)$$

式中,  $\oplus 8$  是模 8 运算。因为方向码的变化是以 8 次为一个周期。可见, 由于图像旋转, 会导致同一轮廓的方向码发生变化, 即缺乏旋转不变性。为了解决该问题, 文献 (王小玲等 2006) 给出了最小和统计方向码的定义, 即

$$D_{\min} = \min \{X_i \cdot D \mid i = 0, 1, \dots, 7\} \quad (4-84)$$

最小和统计方向码通过最小和的约束, 实际上在形状 8 个可能的方向中唯一确定其中某一视角描述对象形状。这时, 统计方向码才可以被用来作为唯一索引。同时, 文中还规定, 如果有多个统计方向码满足上述条件, 则选择沿顺时针方向遇到的第一个方向码为最终表示。因此, 最小统计方向码在一定程度上解决了旋转不变性的问题。同时, 文章也证明最小统计和方向码具有成比例的特点, 这为消除尺度不变性创造了条件。

### 3. 存在问题分析

上述两种方法存在一个共同的问题——都没有考虑方向链码的空间分布特征。设 {000007666665444443222221} 及 {666654443220001222007644} 分别表示两个边缘轮廓的链码串，它们表示的形状如图 4.22 (a) 和 (b) 所示。按照上述方法，它们具有相同的 CCH 及 MSSDC，可是它们所表示的边缘形状却大不相同，主要原因就是上述方法没有考虑方向链码的空间分布特性。

从上面的分析可以看出, 链码的空间分布特征也是边缘形状的一种重要特征。为此, 这里给出了几种新的空间特征描述方法: 链码分布矢量 (Chain Code Distribution Vector, CCDV) (Sun J.D. et al 2006a)、链码相关矢量 (Chain Code Coherence Vector, CCCV) (Sun J.D. et al 2006a)、链码熵 (Chain Code Entropy, CCE)、链码空间分布熵 (Chain Code Spatial Distribution Entropy)、链码相关熵 (Chain Code Relativity Entropy) (Sun J.D. et al 2007a, 孙君顶 2008) 及一种综合特征表示的方法来描述链码串的空间特征。

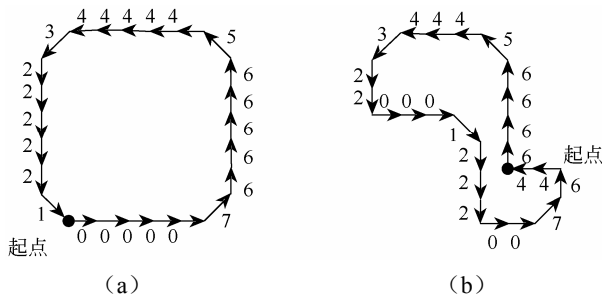


图 4.22 示例轮廓链码表示

### 4.8.2 链码分布矢量 (CCDV)

显然, 链码的空间分布特征体现在链码串中, 主要表现为链码在链码串中位置的变化。为此, 首先提取链码串中链码位置的变化作为链码的空间分布特征。为了描述这种位置的变化, 首先给出一个链码串中链码距离的定义 (Sun J.D. et al 2006)。

**定义 8:** 设  $S$  代表任一轮廓的链码串,  $s_i$  ( $s_i \subset S$ ) 代表链码串中相连的、方向为  $i$  的链码序列, 定义链码序列  $s_i$  在  $S$  中出现的序号为该链码串的距离, 同时规定链码串中起始链码序列的距离为 0。

根据该定义, 上述两链码串的链码距离如图 4.23 所示, 针对  $i$  方向链码, 其距离分布如图 4.24 所示。其中,  $S$  表示链码串的链码,  $j$  表示  $i$  方向链码出现的位置 (从 0 开始计),  $m_j$  表示  $i$  方向链码在  $j$  位置连续出现的次数,  $d_j$  表示  $j$  位置  $i$  方向链码序列的距离。

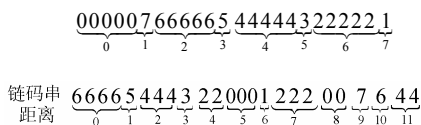


图 4.23 链码距离计算示例

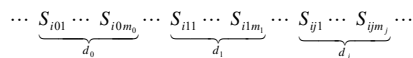


图 4.24  $i$  方向链码距离分布

定义了链码距离后, 可以根据链码距离的变化来反映不同方向链码的空间分布特征, 具体地说, 采用链码距离的均方差来表示, 其定义如下:

$$\sigma_i = \sqrt{\frac{m_0(d_0 - \mu_i)^2 + m_1(d_1 - \mu_i)^2 + \cdots + m_j(d_j - \mu_i)^2 + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}} \quad (4-85)$$

式中,  $\mu_i = \frac{m_0 d_0 + m_1 d_1 + \cdots + m_j d_j + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}$ 。

这里我们在 MSSDC 的基础上来获取链码的空间分布特征, 即首先计算 MSSDC, 然后再获取其空间分布特征。由于 MSSDC 具有旋转和平移不变性,  $\sigma_i$  也具备相应的性质。从链码距离的定义可以看出, 链码距离与起点的选择有关, 因此  $\sigma_i$  也与起点的选择相关。这里首先采用参考文献[7]的方法对链码串进行起点归一化处理, 这样所获取的特征与起点的选择就没有关系。

当某一对象的尺度发生了变化 (如放大或缩小  $\lambda$  倍),  $\sigma_i$  具有尺度不变性。由于尺度的变化仅仅引起链码串长度的变化, 从链码距离的定义可以看出, 尺度的变化对链码距离没有影响, 因此有

$$\mu_i^{(\lambda)} = \frac{(\lambda m_0) d_0 + (\lambda m_1) d_1 + \cdots + (\lambda m_j) d_j + \cdots}{\lambda m_0 + \lambda m_1 + \cdots + \lambda m_j + \cdots} = \mu_i \quad (4-86)$$

$$\begin{aligned} \sigma_i^{(\lambda)} &= \sqrt{\frac{(\lambda m_0)(d_0 - \mu_i^{\lambda})^2 + (\lambda m_1)(d_1 - \mu_i^{\lambda})^2 + \cdots + (\lambda m_j)(d_j - \mu_i^{\lambda})^2 + \cdots}{\lambda m_0 + \lambda m_1 + \cdots + \lambda m_j + \cdots}} \\ &= \sqrt{\frac{m_0(d_0 - \mu_i)^2 + m_1(d_1 - \mu_i)^2 + \cdots + m_j(d_j - \mu_i)^2 + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}} \\ &= \sigma_i \end{aligned} \quad (4-87)$$

可以看出,尺度的变化并没有对 $\sigma_i$ 产生影响。

这里,我们就采用 $\sigma_i$ 作为链码的分布特征,结合链码直方图,给出了 CCDV 的定义如下:

$$\langle (h_0, \sigma_0), (h_1, \sigma_1), \dots, (h_i, \sigma_i), \dots, (h_{n-1}, \sigma_{n-1}) \rangle \quad (4-88)$$

式中, $i$ 表示链码串中的方向数。

针对上面的两个链码串{000007666665444443222221}及{666654443220001222007644},其 CCDV 可分别表示为:  $\langle (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0) \rangle$  及  $\langle (0.21, 1.64), (0.04, 0), (0.21, 1.64), (0.04, 0), (0.21, 4.93), (0.04, 0), (0.21, 4.47), (0.04, 0) \rangle$ 。可以看出,虽然两链码串具有相同的统计特性,但是其空间分布特征却有很大的区别,因此通过其空间分布特征可较好地区分两链码串。

### 4.8.3 链码相关矢量 (CCCV)

针对该方法,将链码串中的链码划分为聚合和非聚合链码两种类型,其定义如下(Sun J.D. 2006a)。

**定义 9:** 在链码串中,针对 $i$ 方向链码,若 $m_j > \tau$  ( $\tau$ 为给定的阈值),则定义链码序列 $S_{ij0}, S_{ij1}, \dots, S_{ijm_j}$ 为聚合链码,否则为非聚合链码。

设 $\alpha_i$ 表示 $i$ 方向聚合链码的统计数目, $\beta_i$ 表示 $i$ 方向非聚合链码的统计数目。则 CCCV 可定义为

$$\langle (\alpha_0, \beta_0), (\alpha_1, \beta_1), \dots, (\alpha_i, \beta_i), \dots, (\alpha_{n-1}, \beta_{n-1}) \rangle \quad (4-89)$$

式中, $i$ 表示链码的方向数。

例如,针对上文链码串{000007666665444443222221}及{666654443220001222007644},如果我们选择 $\tau=4$ ,则其 CCCV 可分别表示为:  $\langle (5, 0), (0, 1), (4, 0), (0, 1), (5, 0), (0, 1), (5, 0), (0, 1) \rangle$  及  $\langle (4, 1), (0, 1), (0, 4), (0, 1), (0, 5), (0, 1), (4, 0), (0, 1) \rangle$ 。可以看出,虽然它们具有相同的统计特征,但其链码相关矢量也存在较大区别。

这里,仍在 MSSDC 的基础上提取 CCCV,因此 CCCV 也具有平移及旋转不变性。很明显,CCCV 与不同方向链码的数目有关,不具备尺度不变性。为了解决这个问题,又对该 CCCV 进行了归一化处理。

设尺度变化为 $\lambda$ ,如果取阈值 $\tau' = \lambda\tau$ ,从而不同尺度下的 CCCV 可表示为  $\langle (\lambda\alpha_0, \lambda\beta_0), (\lambda\alpha_1, \lambda\beta_1), \dots, (\lambda\alpha_n, \lambda\beta_n) \rangle$ , 设 $L_1 = \sum_{i=0}^{n-1} \alpha_i$ ,  $L_2 = \sum_{i=0}^{n-1} \beta_i$ , 经过归一化处理可表示为

$$\left\langle \left( \frac{\alpha_0}{L_1}, \frac{\beta_0}{L_2} \right), \left( \frac{\alpha_1}{L_1}, \frac{\beta_1}{L_2} \right), \dots, \left( \frac{\alpha_i}{L_1}, \frac{\beta_i}{L_2} \right), \dots, \left( \frac{\alpha_{n-1}}{L_1}, \frac{\beta_{n-1}}{L_2} \right) \right\rangle \quad (4-90)$$

这里, $n$ 表示链码串的长度。显然,通过上述处理,CCCV 具有尺度不变性。

### 4.8.4 链码空间分布熵 (CCSDE)

设 $I$ 表示任意轮廓, $I(x, y)$ 表示坐标 $(x, y)$ 处链码的方向,设 $A_i = \{(x, y) | (x, y) \in I, I(x, y) = i, 0 \leq i \leq n\}$ 表示方向为 $i$ 的所有链码的集合,其中 $n$ 表示链码方向数。设 $|A_i|$ 表示集合 $A_i$ 中链码的数目, $C_i = (x_i, y_i)$ 为 $i$ 方向链码的质心。 $x_i$ 和 $y_i$ 定义如下(Sun J.D. et al 2006b, 孙君顶 2008):



$$x_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} x \quad y_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} y \quad (4-91)$$

设  $r_i$  表示  $i$  方向链码同  $C_i$  的最大距离, 其定义为

$$r_i = \max_{(x,y) \in A_i} (\sqrt{(x-x_i)^2 + (y-y_i)^2}) \quad (4-92)$$

对于给定的一个正整数  $M$ , 把  $r_i$  分为  $M$  等份, 然后以质心  $C_i$  为圆心, 以  $(j \times r_i)/M$  为半径 ( $1 \leq j \leq M$ ) 画圆可得到  $M$  个环。那么由内至外每个环和  $A_i$  的交点把  $A_i$  分为  $A_{i1}, A_{i2}, \dots, A_{iM}$ , 我们称  $A_{i1}, A_{i2}, \dots, A_{iM}$  为  $A_i$  的一个分割, 如图 4.25 所示 (以 4-链码为例)。设  $|A_{ij}|$  表示环形区间  $j$  内  $i$  方向链码的数目, 则  $(|A_{i1}|, |A_{i2}|, \dots, |A_{iM}|)$  就构成了  $i$  方向链码的环形分布直方图。

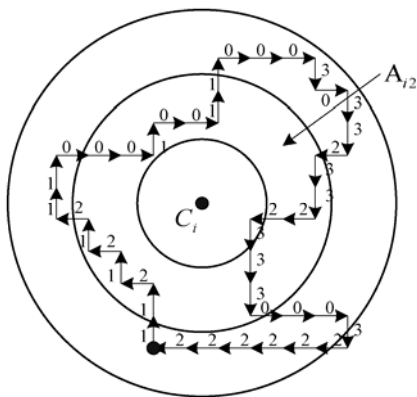


图 4.25 链码环形分布

为了满足尺度不变性, 对环形链码直方图进行了归一化处理, 设归一化后的环形链码直方图为

$$\mathbf{m}_i = (m_{i1}, m_{i2}, \dots, m_{iM}) \quad (4-93)$$

式中,  $p_{ij} = |A_{ij}|/|A_i|$ 。结合信息熵的概念, 则链码分布熵可表示为

$$SE_i(\mathbf{m}_i) = -\sum_{j=1}^M m_{ij} \log_2(m_{ij}) \quad (4-94)$$

链码空间分布熵反映了同方向链码的空间分布特征, 且具有尺度、平移不变性及起点无关性。

结合链码直方图, 新的轮廓特征可以描述为

$$\langle (h_0, SE_0), (h_1, SE_1), \dots, (h_i, SE_i), \dots, (h_{n-1}, SE_{n-1}) \rangle \quad (4-95)$$

#### 4.8.5 链码相关熵 (CCRE)

##### 1. 链码相关熵

针对如图 4.26 所示的轮廓, 其 4-链码可表示为 {10103322}, 从图中可以看出, 该轮廓的主要变化体现在链码  $1 \rightarrow 0, 0 \rightarrow 1, 1 \rightarrow 0, 0 \rightarrow 3, 3 \rightarrow 2, 2 \rightarrow 1$  链码方向的转换。因此, 链码间方向变化也是反映了图像轮廓变化的一个主要特征, 而前面所述的方法都没有考

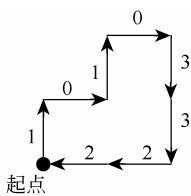


图 4.26 轮廓示例

虑这种特征 (Sun J.D. et al 2007a, 孙君顶 2008)。

如果将链码方向看做不同的状态, 那么链码串就可以看做一状态链, 从而链码方向的变化可以通过状态的变化来体现。很明显, 采用马尔可夫链可以很好地解决该问题, 马尔可夫链的状态转移矩阵就反映了状态间的变化 (林元烈 2002), 这里我们采用一步转移概率矩阵来描述。为了获取更有效的特征表示, 对马尔可夫链的一步转移概率矩阵进行了修改, 新定义的状态转移概率矩阵如下所示:

$$\mathbf{P} = \begin{bmatrix} 0 & p_{01} & \cdots & p_{0(n-1)} \\ p_{10} & 0 & \cdots & p_{1(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{(n-1)1} & p_{(n-1)2} & \cdots & 0 \end{bmatrix} \quad (4-96)$$

式中,  $p_{ij} = \begin{cases} \frac{k_{ij}}{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} k_{ij}}, & i \neq j \\ 0, & i = j \end{cases}$ ,  $k_{ij}$  表示从状态  $i$  转移到状态  $j$  的转移次数,

$\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p_{ij} = 1$ 。在原马尔可夫链的状态转移概率矩阵中,  $p_{ij} = \frac{k_{ij}}{\sum_{j=0}^{n-1} k_{ij}}$  且  $\sum_{j=0}^{n-1} p_{ij} = 1$ 。

在此基础上, 给出了状态  $i$  的相关性直方图定义, 即

$$\mathbf{q}_i = \{p_{i0}, p_{i1}, \cdots, p_{ij}, \cdots, p_{i(n-1)}, p_{0i}, p_{1i}, \cdots, p_{ji}, \cdots, p_{(n-1)i}\}, i \neq j \quad (4-97)$$

该相关性直方图包含两部分, 一是状态  $i$  到其他状态的转移概率, 另一部分是其他状态到状态  $i$  的转移概率。该相关性直方图可以简单的表示为

$$\mathbf{q}_i = \{q_1, q_2, \cdots, q_k, \cdots, q_{2n-2}\} \quad (4-98)$$

式中,  $q_1 = p_{i0}, q_2 = p_{i1}, \cdots, q_{2n-2} = p_{(n-1)i}$ 。可以看出, 如果状态数为  $n$ , 则相关性直方图的维数为  $(2n-2)$ 。结合信息熵的定义, 给出了  $i$  方向链码的相关熵定义, 即

$$RE_i = -\sum_{k=1}^{2n-2} q_k \log_2(q_k) \quad (4-99)$$

$RE_i$  反映了状态  $i$  与其他状态的相关性, 也反映了该方向链码同其他方向链码的相关性,  $RE_i$  越大, 表明相关性越强。结合链码直方图, 新的特征矢量可以描述为

$$\langle (h_0, RE_0), (h_1, RE_1), \cdots, (h_i, RE_i), \cdots, (h_{n-1}, RE_{n-1}) \rangle \quad (4-100)$$

## 2. 链码相关熵特性

很明显, 链码相关熵具有平移不变性, 下面分析其尺度不变性。图 4.27 所示的轮廓与图 4.26 尺度不同, 其链码串表示为 {1100110033332222}。按马尔可夫链理论 (林元烈 2002), 它们的状态转移次数矩阵分别如图 4.28 (a) 和 (b) 所示。可以看出, 由于尺度不同, 它们具有不同的状态转移次数矩阵。如果不考虑同状态间的转移概率, 也就是说, 状态转移次数矩阵中对角元素的值为零, 如图 4.28 (c) 所示, 在这种情况下, 图 4.26 及图 4.27 具有相同的状态转移次数矩阵, 具有尺度不变性。这也是在上面定义状态转移矩阵时规定当  $i = j$  时  $p_{ij} = 0$  的原因, 目的就是为了获取尺度不变性。



$$(e, se_0, se_1, \dots, se_i, \dots, se_{n-1}, re_0, re_1, \dots, re_i, \dots, re_{n-1}) \quad (4-103)$$

## 4.9 基于角点的描述方法

### 4.9.1 轮廓角点提取

基于边界轮廓曲线的角点检测,早在 1975 年, Rosenfeld 等人 (Rosenfeld A. et al 1975) 和 Freeman 等人 (Freeman H. et al 1977) 就提出过用角点强度  $k$  来计算角点。不过,虽然这种方法简单,但是容易受噪声干扰,效果不是很理想。为了将干扰去除,减少边缘毛刺干扰, Asada 等人 (Asada H. et al 1986) 提出首先对边缘进行边缘平滑,即减少了将局部弯曲度突然增大而误判为角点的概率。但角点强度  $k$  是预先确定,还是根据曲线的弯曲度自适应调节,对于检测的结果影响很大。肖茜等人 (肖茜等 2003) 指出自适应的弯曲度测定实际上是要自适应地确定曲线段支持区域的大小,支持区域的选择应该能够根据曲线的弯曲程度自适应地调整,在此支持区域上求取的曲线弯曲度才能较为准确地反映平面对象边界曲线的平滑和弯曲程度。

文献 (钟宝江等 2004) 认为数字化曲线是离散的,是基于像素基础的,这样隐含的一个假设就是数字化曲线上相邻两个像素之间的距离是一个常数。但在实际中该假设并不成立,因此质疑早先对角点的估计方法是否可拟和稳定。在这个发现的基础之上,文章提出了基于曲线加弦长的角点检测方法,主要是在确定支持域时充分考虑相邻像素点之间的实际距离,即相邻的距离应该是 1 和  $\sqrt{2}$ , 并由此出发提出隐式精化数字化曲线的策略,推导出了一种新的角点强度计算公式。利用该公式可以对如尖角和圆角进行区别,检测结果具有旋转不变性。

考虑到角点作为一种重要的信号特征,属于图像细节,按照 Witkin 尺度空间理论,该角点应该在较大的尺度空间存在。滤波器的尺度选择并不是一件容易的事情,要求在滤掉噪声的同时保持边界曲线的基本形状特征。同时曲线上各角点均有着不同尺度的支撑域,无法事先定义出一个最优的分辨率来进行角点检测。在使用多尺度分析后求取不同尺度的空间时,轮廓曲线已经被不同的小波函数所平滑,所以能最大限度地减少边缘毛刺噪声。

Witkin (Witkin A.P 1983) 提出基于尺度空间的图像分析理论后,多尺度曲线分析成为解决该问题的主要方法,在曲线尺度空间中,随着曲线尺度由小变大,一直保持较高弯曲度的点必定是所要求取的角点。基于此, Ratarangsi (Rattarangsi A. et al 1992) 提出基于尺度空间的角点检测思想。Mokhtarian 等人 (Mokhtarian F. et al 1998) 基于尺度空间提出一种新的角点检测算法。他们首先使用 Canny 边缘检测算法从原始图像中提取图像的边缘,角点定义为具有最大绝对曲率的边缘点。在非常小的尺度,由于存在许多噪声,具有最大绝对曲率的边缘点很多,而当尺度增加时,噪声被平滑,只有对应真实角点的最大值保留下来。但是角点的位置也在变化,这时角点的定位很差。所以 Mokhtarian 算法首先在图像中用高尺度检测角点,然后逐步减小尺度,在多个地点尺度出跟踪改善角点定位。因为在降低尺度的过程中只需要计算候选角点,计算量大大降低。

当然,关于轮廓角点提取,还有其他很多有效的方法。本节介绍了四种角点提取方

法,在此基础上,讨论如何有效地提取角点特征并用于基于轮廓的图像检索。

### 4.9.2 基于内角的轮廓角点检测

本节的主要问题是研究轮廓空间特征的提取,因此,针对轮廓角点,仅根据其内角大小简单判断 (Sun J.D. et al 2007b, Wu X.S. et al 2008)。轮廓首先被均匀采样,设采样后的轮廓表示为  $\{s_1, s_2, \dots, s_i, \dots, s_n\}$ ,  $s_i = (x_i, y_i)$  表示  $s_i$  的坐标。为了消除噪声对轮廓造成的影响,采用高斯滤波器分别对  $(x_1, x_2, \dots, x_n)$  及  $(y_1, y_2, \dots, y_n)$  进行平滑处理。

设  $\theta$  表示某一采样点的内角值,若  $\theta > \alpha_1$ , 定义该点为凹角点;若  $\theta < \alpha_2$ , 定义该点为凸角点;若  $\alpha_2 \leq \theta \leq \alpha_1$ , 定义该点为平滑点。 $\alpha_1$  和  $\alpha_2$  为预先设定的阈值,实验表明,  $\alpha_1 \in [1.15\pi, 1.25\pi]$ ,  $\alpha_2 \in [0.7\pi, 0.85\pi]$  时均可取得较好的效果。同时规定:如果轮廓相邻采样点同为凸角点或凹角点,则仅保留这些角点中内角最小或最大的一个,其他角点作为平滑点处理。图 4.30 给出了两幅示例轮廓,图 4.31 给出了示例轮廓角点提取结果,其中“ $\nabla$ ”表示凹角点,“ $\square$ ”表示凸角点,“ $*$ ”表示平滑点。

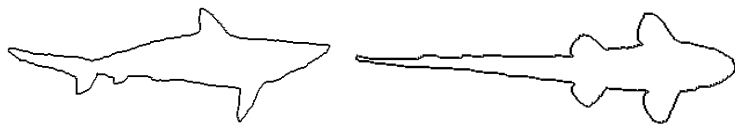


图 4.30 示例轮廓

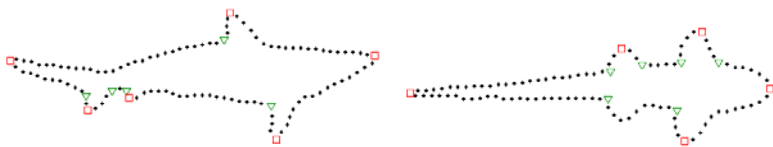


图 4.31 角点提取结果

### 4.9.3 基于链码局部直方图的角点检测

Arrebola 等人通过统计轮廓链码的局部直方图的关系,定位角点的位置和确定角点的凸凹性 (Arrebola F. 1997)。其主要思想是:首先采用链码来描述轮廓,然后对轮廓上每一个点,分别统计其左右两边  $K$  个点内的链码分布信息,并构成两个直方图  $f$  和  $g$ , 然后根据下式得到  $f$  和  $g$  的相关因子系数  $\rho$ , 即

$$\rho = \frac{\sum_i [f(i) - m_f][g(i) - m_g]}{\sqrt{\sum_i [f(i) - m_f]^2 \sum_i [g(i) - m_g]^2}} \quad i \in (0, 1, \dots, 7) \quad (4-104)$$

式中,  $m_f$ 、 $m_g$  分别为直方图  $f$  和  $g$  的均值。可见,若某点的曲率越大,则  $f$  和  $g$  的差异也越大,而因子  $\rho$  却越小。因此,因子  $\rho$  与曲率成反比。将得到的此因子函数通过低通滤波去掉噪声后,角点对应的因子  $\rho$  应是极小值,且因子  $\rho$  应该小于某一个阈值  $\rho_T$ 。实验表明,阈值  $\rho_T$  应在 (0.3, 0.4) 区间,  $K$  的取值应在 5 到 9 之间 (Arrebola F. 1997)。

通过上面的处理,将得到轮廓上每一个点的因子,然而此时并不能区分角点的类别。可通过下式进行修正,修正因子系数为

$$\rho_1 = |\rho - 1| \cdot \text{Sign} \quad (4-105)$$

式中,  $\text{Sign}$  表示当前点的左右直方图走向。通过判别  $\rho_1$  的符号, 可将轮廓点分为平滑点, 凸角点和凹角点。图 4.32 给出了一幅示例轮廓, 图 4.33 给出了  $\rho_1$  的分布, 图 4.34 中给出了轮廓角点提取的初步结果, 其中 “ $\nabla$ ” 表示凹角点, “ $\square$ ” 表示凸角点。



图 4.32 示例轮廓

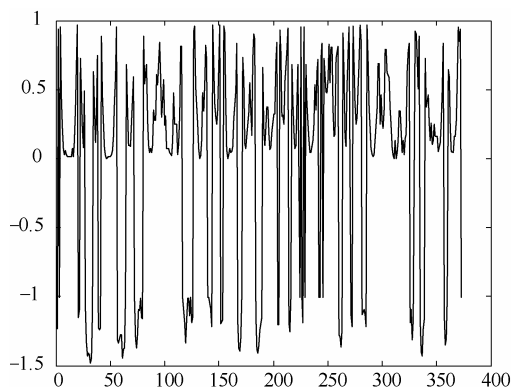


图 4.33  $\rho_1$  的分布

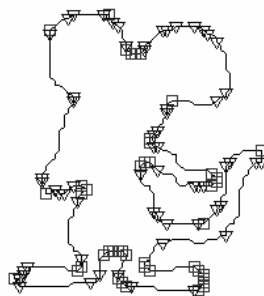


图 4.34 角点提取结果

#### 4.9.4 基于CSS的角点检测

在 4.2.3 节, 介绍了一种成功应用于形状特征描述的算子, 即曲率尺度空间描述符 (CSSD)。同时, 利用曲率尺度空间也可进行角点的检测, Mokhtarian 等提出了一种基于这样的检测方法 (Mokhtarian F. et al 1998)。这种方法的特点是, 在较大的尺度下用曲率公式计算出图像轮廓某点处的曲率, 找出局部极值点, 再通过阈值技术来检测角点, 最后在较小的尺度下对检测出的角点进行定位。

首先找到轮廓上 T 型交叉点, 标记为 T 型角点。然后以高斯函数的参数  $\sigma$  为尺度因子, 在一个较高的尺度上计算轮廓曲线上任意一点处的曲率为

$$k(u, \sigma) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{(X_u(u, \sigma)^2 + Y_u(u, \sigma)^2)^{3/2}} \quad (4-106)$$

式中,

$$X_u = x(u) \otimes g_u(u, \sigma) \quad X_{uu} = x(u) \otimes g_{uu}(u, \sigma) \quad (4-107)$$

$$Y_u = y(u) \otimes g_u(u, \sigma) \quad Y_{uu} = y(u) \otimes g_{uu}(u, \sigma) \quad (4-108)$$

$\otimes$  是一个卷积符号,  $u$  为弧长参数,  $g(u, \sigma)$  为高斯函数,  $g_u(u, \sigma)$  和  $g_{uu}(u, \sigma)$  是对  $g(u, \sigma)$  分别对  $u$  求一阶和二阶导数; 然后把局部曲率最大点作为候选角点, 如果某个候选角点处的曲率值大于阈值  $t$ , 并且大约是相邻局部曲率最小点处曲率值的 2 倍, 则把该角点作为正确角点; 最后在较低的尺度下定位这些角点, 并和 T 型角点进行比较, 剔除相隔较近的两个中的一个角点。

CSS 算法获得了比较好的角点检测结果, 并且对噪声也不很敏感, 但是该算法仍有三方面问题: 其一, 在检测角点的过程中使用是单一的高尺度, 这样很容易漏掉一部分正确的角点; 其二, 当尺度过大时不能检测到真正的角点; 第三, 用于决定角点取舍的全局阈值  $t$  是预先固定的, 它的取值对最终的结果影响很大。因此, Mokhtarian 等人对轮廓曲线长度不同的图像选用不同的尺度来减少漏掉角点的情况 (Mokhtarian F. et al 2001)。然而合适的尺度并不能仅仅由图像的轮廓曲线长度所决定, 并且轮廓曲线的长度也受边缘检测算法的影响。X.C.He 等人提出了 ACSS 算法 (He X.C. et al 2004), 是先在低的尺度下, 计算曲率获取初始的候选角点, 然后利用取局部阈值和角度估计的方法移除圆角点和量化噪声和边缘细节产生的错角点, 但是引入了过多的阈值参数, 使得控制的难度和计算的复杂度都比较大。针对图 4.32 的示例轮廓, 图 4.35 给出了采用 CSS 和 ACSS 检测方法提取示例轮廓角点的结果, 其中 “□” 表示检测出的角点。

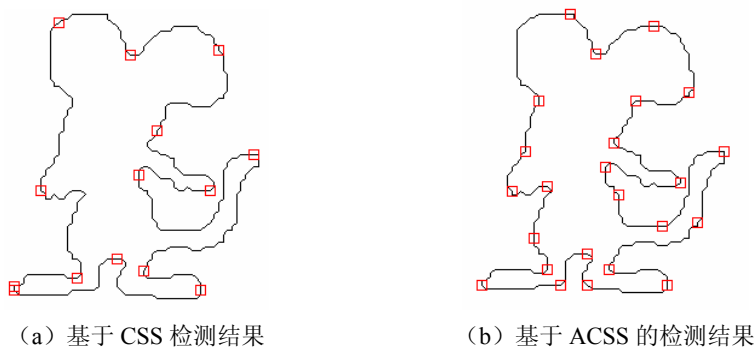


图 4.35 两种角点检测结果

在基于尺度空间的特征检测中, 融合多尺度信息的方法有多种, 一种是大尺度检测小尺度跟踪定位, 如上所述的 CSS 方法; 另一种就是, 利用多尺度乘积来增强特征信息, 同时抑制噪声对特征的影响。张小洪等人则是根据多尺度积的思想, 提出了一种增强角点曲率的方法 MSCP (张小洪等 2007)。即根据式 (4-106) 就可以计算出在第  $j$  尺度下的曲率  $k(u, \sigma_j)$ , 其多尺度曲率积为

$$P_N(u) = \prod_{j=1}^N k(u, \sigma_j) \quad (4-109)$$

利用这种方法, 取曲率的乘积大于某个阈值  $k$  的局部极大值点作为角点。图 4.36 给出了采用 MSCP 思想对图 4.32 示例轮廓在不同尺度空间下的曲率乘积的曲线图。根据  $N$  取 1、2、3、4、5 时曲率乘积的分布, 取适当阈值  $|k| < 0.0004$ , 可检测出示例轮廓的角

点,如图 4.37 所示。

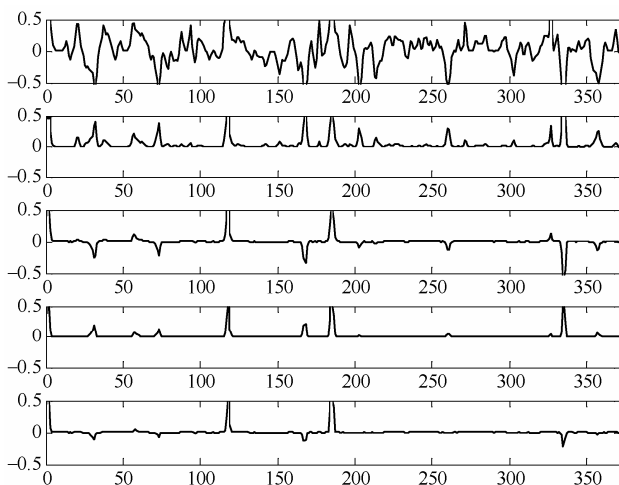


图 4.36 当  $N$  分别为 1、2、3、4、5 时的曲率乘积

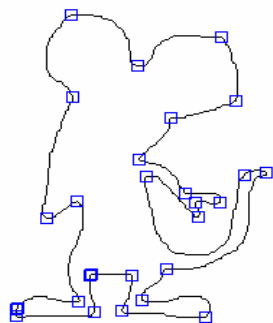


图 4.37 基于 MSCP 的检测结果

#### 4.9.5 改进的多尺度角点检测方法

前面介绍了几种轮廓角点提取方法,虽然都能够在某种状态下达到很好的效果,但也都存在着不足。其中,基于内角的方法简单易于实现,然而因其是等步长取点,虽然能减少噪声的影响,但却可能会漏掉一些关键点,若在此基础上再高斯平滑,则得到的点集会失去很多轮廓细节,因此只能局限于轮廓简单且对检测精度要求不高的物体。基于链码的算法利用轮廓链码分布关系,提取弯曲因子和方向,但受噪声和琐碎细节的干扰,只能定位到角点的范围,很难准确定位和剔除干扰。可见,在单一尺度下很难摆脱噪声和冗余细节的干扰并达到很好的检测效果。多尺度积方法将多尺度下的特征融合在一起,从而增强特征信息,同时抑制噪声的干扰。随着尺度的增大,噪声和冗余被逐步减弱,因此将同一点在所有尺度下的曲率相乘,能够将噪声和冗余对应的值越来越接近 0,而曲率极大值对应的点逐渐锐化,从而可以阈值化得到角点。但是,由于角点的出现只是针对一个或者一些尺度空间,这就意味着这类角点针对小的尺度来说是角点,而在大尺度下将被作为假角点滤除。虽然此方法能够在很大程度上增强了一些正确的角点,然而却平滑掉了很多角点。

因此,在充分利用多尺度信息和角点的尺度不确定性,本节给出了一种基于多尺度的曲率多项式检测方法 (Multi-Scale Curvature Polynomial, MCP)。此方法在增强角点极大值,抑制噪声和冗余的同时,也增强了部分正确角点对应的曲率局部极大值,同时考虑了角点不同的特性,将检测角点的区别分为凸角点和凹角点。

设  $g(u, \sigma_j)$  是高斯函数  $g(u)$  在不同尺度  $\sigma_j$  下的形式,可表示为

$$g(u, \sigma_j) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_j^2}} \quad (j=1, 2, \dots) \quad (4-110)$$

根据前述方法,可获得不同尺度下的轮廓和轮廓上各点对应的曲率值。不同的尺度对曲率的作用程度不同,对不同角点的作用程度也不同,分别对不同尺度下局部极大值点对应的曲率采用加权和,而对非极值点采用曲率积的形式,达到增强角点的同时平滑噪声和冗余



细节。可将第  $m$  尺度下轮廓上某点对应的曲率  $E_N(u)$  表示为

$$E_N^{(j)}(u) = \begin{cases} E_N^{(j-1)}(u) + k(u, \sigma_j), & j \text{ 尺度下 } u \text{ 处为局部极值} \\ E_N^{(j-1)}(u) * k(u, \sigma_j), & j \text{ 尺度下 } u \text{ 处不是局部极值} \end{cases} \quad (4-111)$$

式中,  $j=2,3,\dots,N$ ,  $E_N^{(1)}(u) = k(u, \sigma_1)$  可以在连续的一些尺度空间上根据此式计算出每点的曲率多项式特征。在计算出轮廓上所有点的曲率多项式值后, 能够通过取阈值的方法提取出角点。由式 (4-111) 可知, 计算出的曲率多项式值是正负可分的, 分别代表着轮廓的凸凹性。因此, 可根据此判断提取出的角点的凸凹性。又因为不同图像的轮廓体现的凸凹性不同, 在取阈值提取角点时, 应对正的局部极大值和负的局部极小值分别对待。根据实验获知, 正阈值  $E_+$  取值范围为  $(0.1 \sim 0.5)$ , 负阈值  $E_-$  取值范围为  $(-1 \sim -0.1)$ 。对图 4.32 的示例轮廓图进行检测, 图 4.38 列出了当  $N$  分别为 1、2、3、4、5 时的多项式曲率分布。图 4.39 给出了当正阈值  $E_+$  取 0.12, 负阈值  $E_-$  为 -0.12 时检测的结果, 其中 “ $\nabla$ ” 表示凹角点, “ $\square$ ” 表示凸角点。

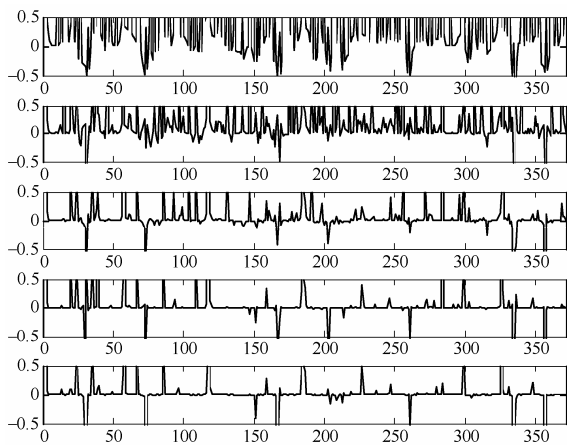


图 4.38 当  $N$  分别为 1、2、3、4、5 时的多项式曲率分布

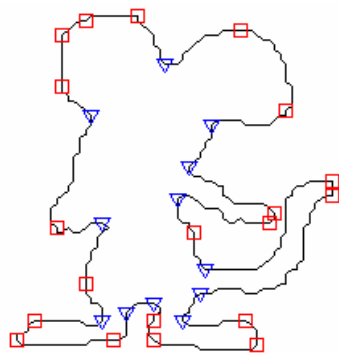


图 4.39 基于曲率多项式的检测结果

#### 4.9.6 角点检测算法比较

Mokhtarian 以角点的数量和位置评价角点检测器的稳定性, 并提出了相应的角点数一致性 (Consistence of Corner Numbers, CCN) 概念 (Farahnaz M. et al 2001), 通过旋转、尺度等变换验证检测器的稳定性。设原图像中角点的数目为  $N_0$ , 变换后检测器检测出的角点数目为  $N_d$ , 则角点数一致性可以表示为

$$CCN = 100\% * 1.1^{-|N_d - N_0|} \quad (4-112)$$

可见, 当 CCN 的值越大, 角点检测器稳定性能越好。

Sarkar 提出了一种品质因数 (Figure of Merit, FOM) 的计算方法 (Sarkar D. 1993)。计算出压缩比 (Compression Ratio, CR), 然后比上误差平方积分 (Integral Square Error, ISE) 所得, 可表示为

$$\text{FOM} = \frac{\text{CR}}{\text{ISE}} = \frac{\frac{N}{N_d}}{\sum_{i=1}^N d_i^2} \quad (4-113)$$

式中,  $N$  为轮廓上所有点数目,  $d_i$  为轮廓与以检测出角点拟和出的多边形之间的距离。

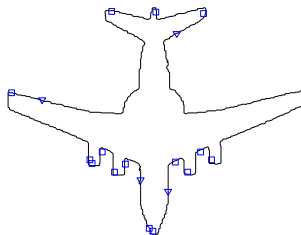
为了从准确性、稳定性和计算代价等各方面评价算法性能, 随机选择了 50 幅不同语义的图像, 然后分别对其进行旋转变换, 并统计角点数目计算出平均的 CCN 值和计算时间, 并提取一组 ISE 和 FOM 值, 如表 4.16 所示。其中列出了基于内角 (Intern-angle)、基于链码局部直方图 (Chain)、基于曲率尺度空间 (CSS)、基于改进的曲率尺度空间 (ACSS)、基于多尺度曲率积 (MSCP) 和基于多尺度曲率多项式 (MCP) 等检测方法计算出的参数值。从表中可知, 基于多尺度曲率多项式 (MCP) 的检测器虽然在计算时间上略劣于 ACSS, 但 ISE、FOM 和平均 CCN 均稍高于其他检测器。图 4.40 给出了将六种检测方法关于 airplane 图像的六种角点检测图。其中 “▽” 表示凹角点, “□” 表示凸角点。可以看出, MCP 取得最优的效果。

表 4.16 实验结果

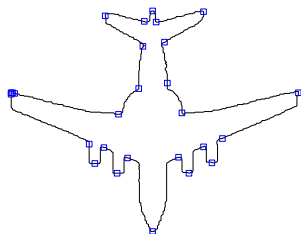
方 法 参 数	基 于 内 角	基 于 链 码	曲率尺度空间	改进曲率 尺度空间	多尺度曲率积	多尺度曲率 多项式
误差平方积分	4064.04	4230.74	3859.62	3620.15	4008.2	3582.68
品质因数	0.038	0.031	0.041	0.044	0.039	0.048
平均角点数一致性	41	44	52	70	64	75
平均检测时间	0.7674	1.3168	1.8039	0.3774	0.6988	0.5461



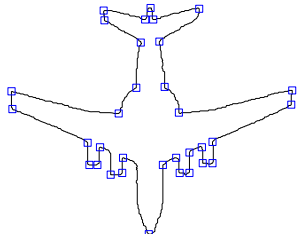
(a) 原始图像



(b) 基于内角的检测结果



(c) 曲率尺度空间的检测结果



(d) 改进曲率尺度空间的检测结果

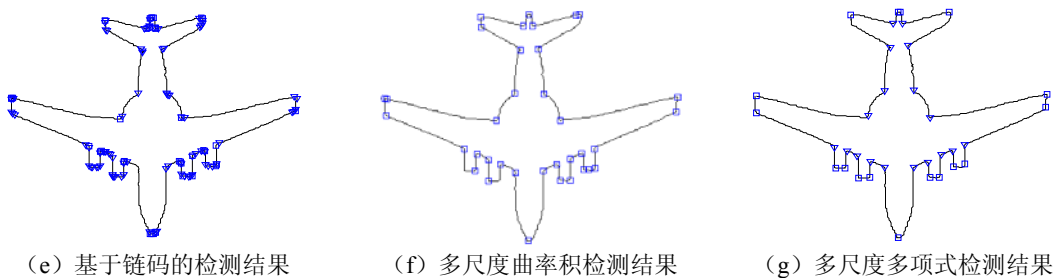


图 4.40 airplane 图像角点检测结果

### 4.9.7 距离直方图

从 4.8.1 节到 4.8.6 节，我们介绍了一些角点提取算法，从本节开始，将介绍几种轮廓角点的特征描述方法。

距离直方图定义为 (Sajjanhar A. 2003): 计算轮廓点同其质心间的距离并将其量化到一定的区间, 设量化等级为  $L$ , 量化到每一区间的轮廓点数为  $d_i (i=0,1,\dots,L-1)$ , 则距离直方图表示为

$$D = (d_0, d_1, \dots, d_{L-1}) \quad (4-114)$$

显然, 上述距离直方图具有平移及旋转不变性, 但不具备尺度不变性。在上述距离直方图的基础上, 针对轮廓角点及平滑点, 分别给出了其距离直方图。同时, 为了获取尺度不变性, 对其进行归一化处理, 设归一化后的距离直方图表示为

$$H^{(k)} = (h_0^{(k)}, h_1^{(k)}, \dots, h_{L-1}^{(k)}) \quad (4-115)$$

式中,  $h_i^{(k)} = c_i^{(k)} / n_k$ ,  $n_k$  ( $k=1,2,3$ ) 分别表示凸角点、凹角点及平滑点的数目,  $c_i^{(k)}$  ( $k=1,2,3$ ) 表示量化到区间  $i$  的凸角点、凹角点及平滑点的数目。

然而, 距离直方图仅考虑了轮廓点同其质心间的距离分布统计特性, 并未考虑角点相互间的分布特性以及平滑点的分布特性。为了解决该问题, 文中提出采用相对位置分布及相关单元熵来进一步描述角点及平滑点的特征。

### 4.9.8 相对位置分布

从图 4.41 可以看出, 同类角点 (凸角点或凹角点) 间的距离关系也是反映角点分布的一个重要特征, 这里就采用角点间的这种距离关系来进一步描述角点特征。

设  $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$  表示某轮廓所有凸角点的集合,  $a_i = (x_i, y_i)$  表示  $a_i$  的坐标,  $m$  表示凸角点的数目。设  $d_{ij}$  表示相邻两凸角点间的欧氏距离, 其中  $i=1,2,\dots,m$ ,  $j=(i+1) \bmod m$ , 取距离均方差  $\delta$  作为相对位置分布特征。同理, 可以得到凹角点的空间位置分布特征。

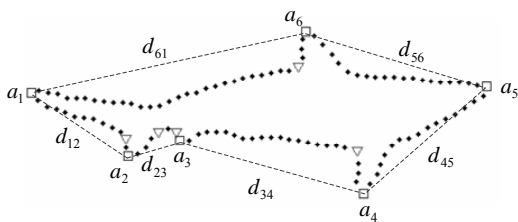


图 4.41 相对位置分布

显然, 相对位置分布特征  $\delta$  具有旋转及平移不变性, 为了获取尺度不变性, 文中对其进行了归一化处理, 从而将相对位置分布定义为

$$\text{RAD} = \delta / d_{\max} \quad (4-116)$$

式中,  $d_{\max} = \max \{d_{ij}\}$ 。

#### 4.9.9 相关单元熵

从图 4.41 亦可看出, 轮廓凹凸点的变化反映了轮廓的主要特征, 但另一方面平滑点也是反映轮廓特征的重要因素。在 4.8.3 节, 采样距离直方图描述平滑点与轮廓质心的距离分布关系, 为了进一步表示平滑点特征, 这里又提出采样相关单元熵进一步描述平滑点特征。

设  $M$  表示轮廓角点的数目 (包括凸角点和凹角点), 这里将相邻两角点间的平滑点称为一个单元, 单元内平滑点的数目用  $l_i$  表示,  $i=1, 2, \dots, M$ , 轮廓所有平滑点的数目用  $K$  表示。结合空间分布熵的定义 (Sun J.D. et al, 2006), 给出了相关单元熵的定义, 即

$$\text{RUE} = -\sum_{i=1}^M p_i \log_2(p_i) \quad (4-117)$$

式中,  $p_i = l_i / K, i=1, 2, \dots, M$ 。

### 4.10 基于矩的轮廓描述方法

目前, 矩技术已经广泛的应用到图像处理、计算机视觉及模式识别技术中 (Hu 1962, Wong et al 1995, Chima et al 1999, Annupan R. 2007), 如景物匹配、图像重建、图像压缩、对称性检测、图像规格化、纹理分割、边缘检测、目标识别和图像检索等。并且许多矩的快速算法也被提了出来 (Wu, C.H. et al. 2007, Liu, J. 2004)。4.3 节讨论了常用的基于区域的矩描述方法, 本节主要讨论基于边缘的矩描述方法。

#### 4.10.1 轮廓矩

物体的轮廓是描绘物体形状的重要特征, 可以将其看成是一条具有方向性的相互衔接的链条。在利用不变矩对其进行描述时, 应首先研究下各个不变矩的含义和物理体现, 并结合边界轮廓的特性, 使其更充分的描述轮廓特征。

利用几何矩进行基于区域的特征描述时, 不同的阶数表示了不同的特性。文献 (Prokop R. J. et al 1992) 中给出了各阶几何矩的含义。针对轮廓矩来说, 各阶几何矩的物理意义可以表述如下。

## (1) 零阶几何矩

根据几何矩的定义, 图像轮廓的零阶几何矩  $M_{00}$  定义为

$$M_{00} = \iint_c dx dy \quad (4-118)$$

可见, 对于边界轮廓, 几何矩  $M_{00}$  表示边界点数目的总和, 即周长。

## (2) 一阶几何矩

一阶几何矩  $M_{01}$  及  $M_{10}$  是图像关于  $x$  轴和  $y$  轴的矩, 可以用来确定边界轮廓的几何中心点, 其坐标  $(\bar{x}_0, \bar{y}_0)$  可以表示为

$$\bar{x}_0 = \frac{M_{10}}{M_{00}}, \quad \bar{y}_0 = \frac{M_{01}}{M_{00}} \quad (4-119)$$

通常, 将坐标系原点移至轮廓形状的几何中心点的矩, 称为中心几何矩。这一变化使矩的计算独立于图像的坐标系。其中, 中心几何矩  $\mu_{pq}$  可表示为

$$\mu_{pq} = \iint (x - \bar{x}_0)^p (y - \bar{y}_0)^q dx dy \quad (4-120)$$

## (3) 二阶中心几何矩

由式(4-104)可以得出形状轮廓的二阶和三阶矩, 其中二阶矩包括  $\mu_{11}$ 、 $\mu_{20}$  和  $\mu_{02}$ 。 $\mu_{11}$  表示轮廓的倾斜度, 若  $\mu_{11}$  大于零则表示图像向左上倾斜, 若小于零则表示向右上倾斜;  $\mu_{20}$  和  $\mu_{02}$  分别表示轮廓在水平和垂直方向上的伸展度, 若  $\mu_{20}$  大于零则表示图像下部的水平伸展比图像上部大, 小于零则是上部的水平伸展比下部大; 若  $\mu_{02}$  大于零则表示图像右边的垂直伸展比左边大, 小于零则左边的垂直伸展比右边大。另外,  $\mu_{20}$  和  $\mu_{02}$  为对  $x$  轴和  $y$  轴的惯性矩, 通过组合它们可以确定几个重要的特性, 如主轴比和方向、椭圆性等。

## (4) 三阶矩或三阶以上矩

三阶或三阶以上矩, 则是轮廓细节和低阶矩变化的具体表现。其中,  $\mu_{30}$  和  $\mu_{03}$  描述了边界曲线投影的扭曲程度, 即关于均值对称分布的偏差程度。分别表示了轮廓在水平和垂直方向上的重心偏移度,  $\mu_{30}$  大于零表示重心偏左, 小于零则是偏右;  $\mu_{03}$  大于零表示重心向上偏移, 小于零则表示重心向下偏移;  $\mu_{21}$  和  $\mu_{12}$  表示轮廓的水平与垂直伸展的均衡程度, 其中  $\mu_{21}$  大于零表示轮廓上部的水平伸展比下部大, 小于零则表示下部比上部的水平伸展程度大;  $\mu_{12}$  大于零表示轮廓右边的垂直伸展比左边大, 小于零则表示轮廓左边的垂直伸展比右边大。

## 4.10.2 Chen不变矩

自从 Hu 提出不变矩以后, 不变矩在形状识别和分类中获得了广泛的应用。但不变矩是针对区域像素计算的, 它不能用于边界的检测。为使其可用于边界形状的检测, Chen (Chen C.C.1993) 在区域不变矩的基础上, 提出了矩的轮廓描述, 其形式表示为

$$M_{pq} = \int_C x^p y^q ds \quad (4-121)$$

式中,  $p, q = 0, 1, 2, 3, \dots$ ,  $\int_C$  表示曲线  $C$  积分,  $ds = \sqrt{(dx)^2 + (dy)^2}$ 。对于数字轮廓, 轮廓矩可表示为

$$M_{pq} = \sum_{(x,y) \in C} x^p y^q \quad (4-122)$$

轮廓中心矩定义为

$$\mu_{pq} = \sum_{(x,y) \in C} (x - \bar{x})^p (y - \bar{y})^q \quad (4-123)$$

式中,  $\bar{x} = M_{10}/M_{00}$ ,  $\bar{y} = M_{01}/M_{00}$  上述不变矩具有旋转不变性, 为了获取尺度不变性, 归一化的中心矩可以表示为

$$\eta_{pq} = \mu_{pq} / \mu_{00}^\gamma \quad (4-124)$$

式中,  $\gamma = p + q + 1$ 。轮廓矩的 7 个不变量的计算类似 Hu 不变矩的 7 个不变量, 这里不再详细给出。

目前, 轮廓矩已经在模式识别方面得到了广泛的应用 (Chima Y.C. et al 1999, Liu Y.S. et al 2004)。

### 4.10.3 边界序列矩

边界序列矩 (Contour Sequence Moments) 由 Gupta and Srinath (Gupta L. et al 1987) 提出。一般物体的轮廓能够通过分段的多项式拟合、链码、梯度角描述, 那么轮廓就可以表示为一个有序的向量几何。此描述方法先假设边界点为  $\{(x(i), y(i)), i = 1, 2, \dots, N\}$ , 边界点的数目为  $N$ , 由边界点围成的中心点即为  $(\bar{x}, \bar{y})$ , 则可将每点与中心点的距离  $z$  作为此边界的一维描述函数  $z(i), i = 1, 2, \dots, N$ 。图 4.42 给出了两幅鱼类示例图, 图 4.43 给出了它们对应的  $z(i)$  函数。



图 4.42 两幅鱼类的示例轮廓

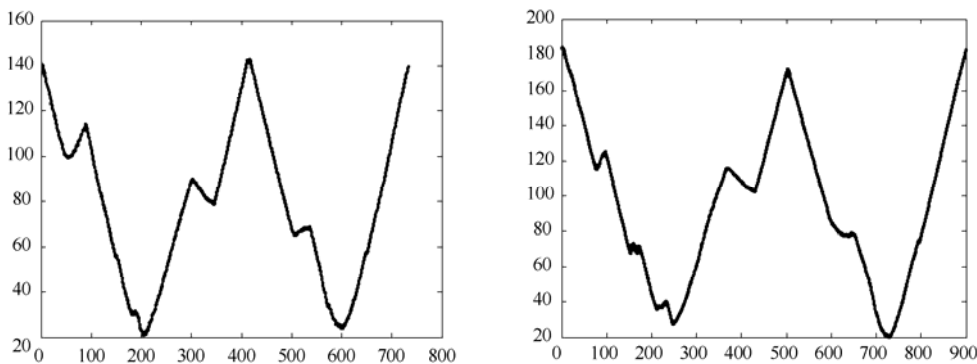


图 4.43 对应的  $z(i)$  函数曲线图

利用此一维描述函数, 可以获取出边界轮廓的几何矩, 即

$$M_p = \frac{1}{N} \sum_{i=1}^N [z(i)]^p \quad (p=1, 2, \dots) \quad (4-125)$$

$p$  阶中心矩为

$$\mu_p = \frac{1}{N} \sum_{i=1}^N [z(i) - M_1]^p \quad (4-126)$$

选取 4 个低阶矩的组合, 可用于目标的识别与分类, 且这 4 个低阶矩具有尺度、旋转和平移不变性 (Gupta L. 1987)。其形式可表示为

$$S_1 = \mu_2 / \mu_1^2, S_2 = \mu_3 / (\mu_2)^{3/2} \quad (4-127)$$

$S_1$  称为归一化的幅度变化量, 它是非负数。当且仅当边界为一圆时  $S_1 = 0$ ;  $S_2$  为与形状的对称程度有关的量, 称为歪斜度  $\mu_3 / (\mu_2)^{3/2}$ ; 与密度函数的峰值峭度有关的量, 称为峭度。实验证明, 利用这 4 个参数可实现对形状的可靠识别和分类。

#### 4.10.4 极半径不变矩

极半径不变矩由曹茂永等在文献中提出 (曹茂永等 2004), 这里主要介绍其对轮廓曲线的描述矩。设  $p(x, y)$  为边界上的任一点,  $r$  为该点到形心  $(\bar{x}, \bar{y})$  的距离。那么, 第  $p$  阶极半径矩和中心矩的离散形式定义为

$$m_{np} = \frac{1}{N} \int \left( \frac{r}{\bar{r}} \right)^p ds \quad (4-128)$$

式中,  $N$  为边界的周长,  $\bar{r} = \frac{1}{N} \int r ds$  为平均极半径,  $ds = r d\theta$  为极坐标下的线积分元; 当边界的形状均匀缩放  $\alpha$  倍时, 周长变为  $\alpha N$ 。可见, 此半径矩和中心矩具有平移、旋转和尺度不变性。 $m_{np}$  和  $m_{ncp}$  为不变量, 那么它们的组合也必是不变量, 选取适当的组合可用于物体的识别与匹配。从中选取 5 个低阶矩的组合, 可用于目标的识别与分类, 且这 5 个低阶矩具有尺度、旋转和平移不变性。对于离散的数字轮廓, 用求和代替积分, 设取轮廓上所有的点  $N$  为采样点, 则 5 个不变矩可表示为

$$V_1 = m_{n2} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i}{\bar{r}} \right)^2 \quad (4-129)$$

$$V_2 = m_{nc2} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^2 \quad (4-130)$$

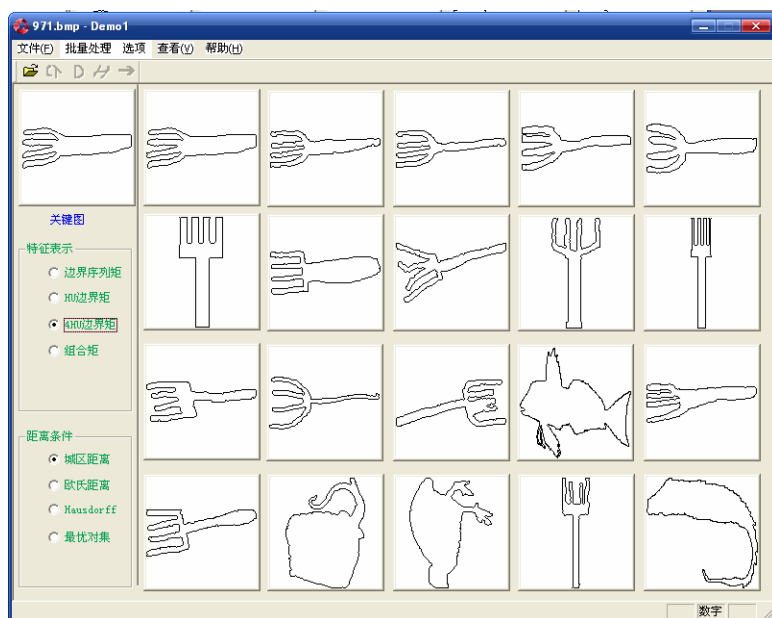
$$V_3 = m_{nc3} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^3 \quad (4-131)$$

$$V_4 = m_{nc4} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^4 \quad (4-132)$$

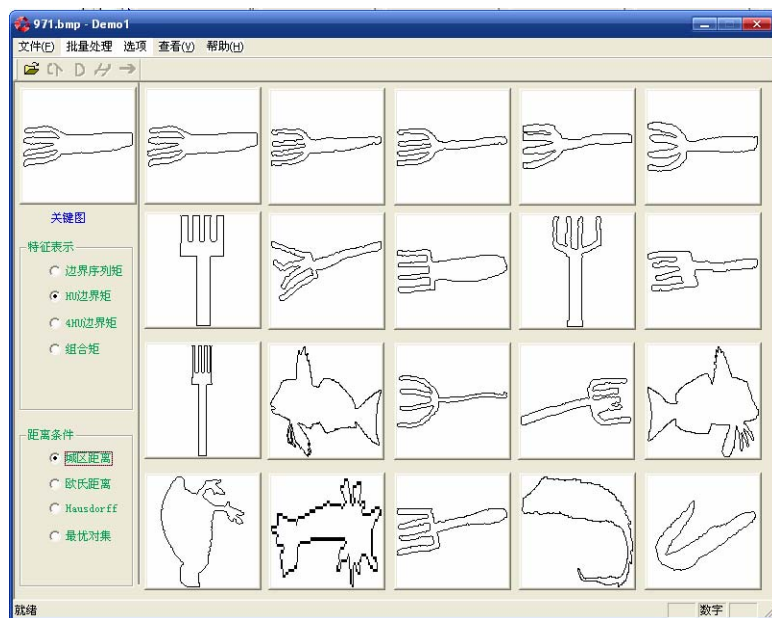
$$V_5 = \frac{m_{nc2}}{\sqrt{m_{nc4}}} = \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^N (r_i - \bar{r})^2}{\sqrt{\sum_{i=1}^N (r_i - \bar{r})^4}} \quad (4-133)$$

#### 4.10.5 组合矩

对于图像检索来说, 寻求的是一种相似性匹配。而对于高阶矩来说反映的轮廓的细节, 因此, 采用高阶矩进行检索并不能有效的提高图像的检索效率。如图 4.44 所示, 图 (a) 是采用 Chen 不变矩的前 4 阶矩组成特征矢量进行检索的结果, 图 (b) 是采用 7 个矩进行检索的结果。



(a) 前4阶矩检索结果



(b) 前7阶矩检索结果

图 4.44 Chen 不变矩检索结果

同时在边界序列矩的基础上，导出了三个具有尺度、旋转及平移不变性的三个矩，表示如下 (Sun J.D. et al 2008):

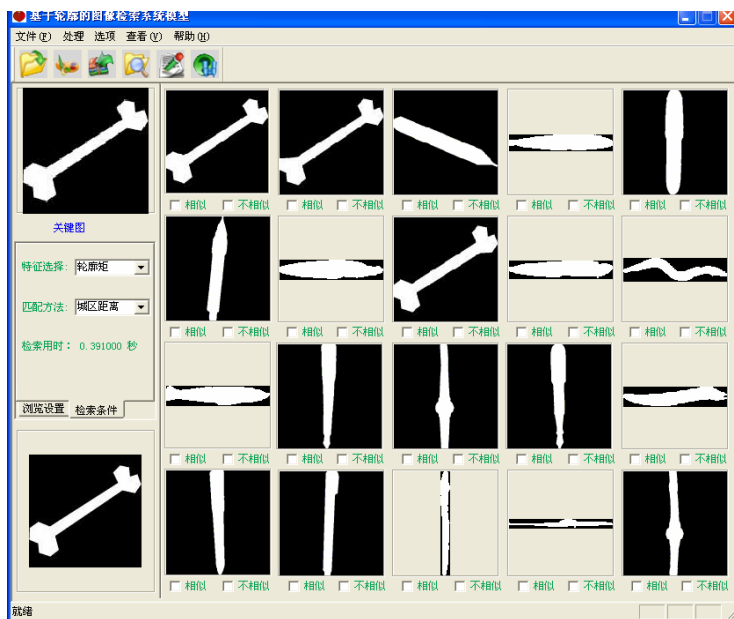
$$\begin{aligned} T_1 &= \mu_2 / (M_1)^2 \\ T_2 &= \mu_3 / (M_1)^3 \\ T_3 &= \mu_4 / (M_1)^4 \end{aligned} \quad (4-134)$$



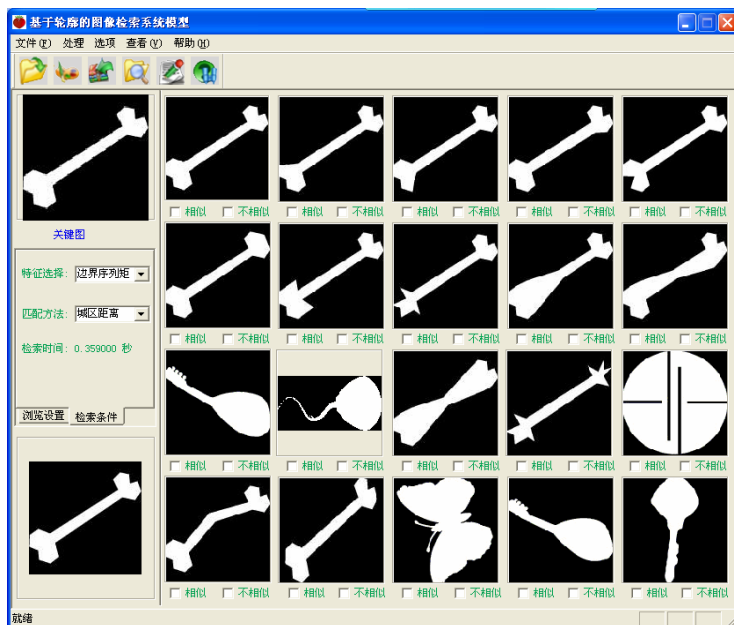
结合 Chen 不变矩的前 4 阶矩, 给出了一个组合矩的定义 (Sun J.D. et al 2008), 即

$$C = [\phi_1, \phi_2, \phi_3, \phi_4, T_1, T_2, T_3] \quad (4-135)$$

为了测试上述矩的效果, 组建了实验图像库, 包含 MPEG-7 标准形状检索库在内的共 2075 多幅图像, 这些图像分为 73 个语义类。其中包括: 鸟类、家禽、鱼类等动物, 也包含有树木、花朵、钢笔等常见物体。图 4.45 给出了上述四种矩的一次检索结果比较, 可以看出组合矩取得了最好的检索效果。

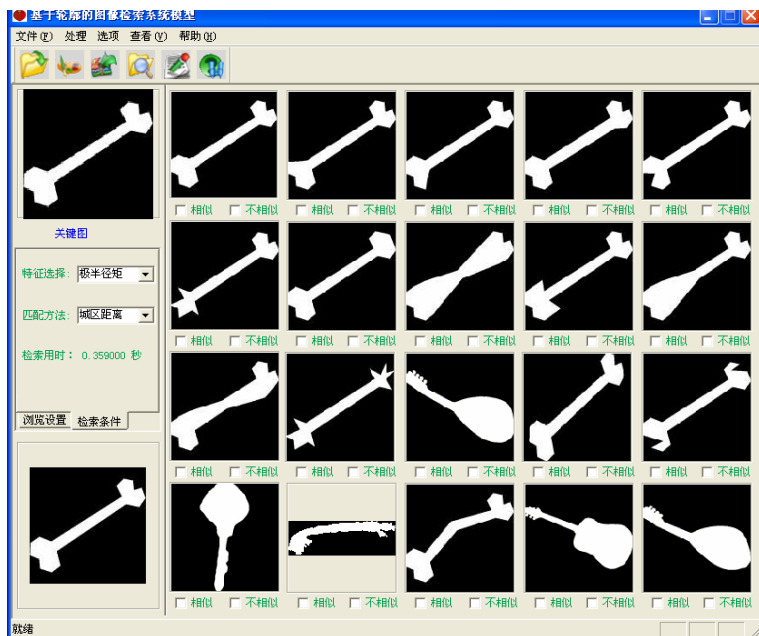


(a) Chen 矩方法

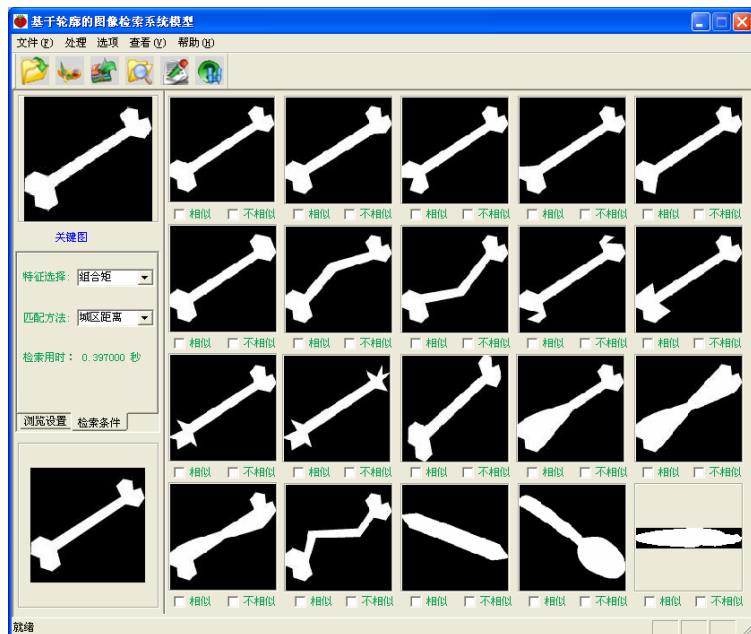


(b) 边界序列矩方法

图 4.45 四种方法检索效果对比



(c) 极半径不变矩方法



(d) 组合矩方法

图 4.45 四种方法检索效果对比 (续)

## 参 考 文 献

- [1] Abbasi S., Mokhtarian F., Kittler J., 1999. Curvature scale space image in shape similarity retrieval, *Multimedia Systems* 7: 467-476.
- [2] Abbasi S., Mokhtarian F., Kittler J., 2000. Enhancing CSS-based shape retrieval for objects with shallow concavities, *Image Vision Comput.* 18: 199-211.
- [3] Alajlan N., Elrube I., Kamel M.S. et al, 2007. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*, 40(7): 1911-1920.
- [4] Ang Y.H., Zhao Li, Ong S.H., 1995. Image retrieval based on multidimensional feature properties. In: *Proc of SPIE: Storage and Retrieval for Image and Video Database*. San Jose, CA, 2420: 47-57.
- [5] Annupan, R., Stanislav, S., 2007. A filter bank method to construct rotationally invariant moments for pattern recognition, *Pattern Recognition Letters* 28, 1492-1500.
- [6] Arbter K., Snyder W.E., Burkhardt H., Hirzinger G., 1990. Application of affine-invariant Fourier descriptors to recognition of 3-D objects, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (7): 640-647.
- [7] Arica N., Yarman Vural F.T., 2003. BAS: a perceptual shape descriptor based on the beam angle statistics, *Pattern Recognition Letters*, 24(9): 1627-1639.
- [8] Arrebola F., Bandera A., Camacho P., et al, 1997. Corner detection by local histograms of contour chain code. *Electronics letters*, 33(21): 1769-1771.
- [9] Asada H., Brady M., 1986. The curvature primal sketch. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 8(1): 2-14.
- [10] Belongie S., Malik J., Puzicha J., 2001. Matching shapes, in: *Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV2001)*, Vol. I, Vancouver, Canada, pp. 454-461.
- [11] Bober M., 2001. Mpeg-7 visual shape descriptor. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6): 716-719.
- [12] Bober M., Kim J.D., Kim H. K., et al, 1999. Summary of the results in shape descriptor core experiment. *ISO/IEC JTC1/SC29/WG11/MPEG99/M4869*.
- [13] Bribiesca E., 1999. A new chain code. *Pattern Recognition* 32, 235-251.
- [14] Bribiesca E. 2000. A chain code for representing 3D curves. *Pattern Recognition* 33: 755-765.
- [15] 曹茂永, 孙农亮, 郁道银. 用于模式识别的极半径不变矩. *计算机学报*, 2004, 6(27): 860-864.
- [16] Chakrabarti K., Binderberger M.O., Porkaew K., et al, 2000. Similar Shape Retrieval in MARs, *IEEE Int. Conf. On Multimedia and Expo (II)*. USA, 709-712.
- [17] Chellappa R., Bagdazian R., 1984. Fourier coding of image boundaries, *IEEE Trans. Pattern Anal. Mach. Intell.* 6(1): 102-105.
- [18] Chen, C.C., 1993. Improved moment invariants for shape discrimination, *Pattern Recognition*, 26(5): 683-686.
- [19] Chetverikov D., Khenokh Y., 1999. Matching for shape defect detection, *Lecture Notes in Computer Science*, Vol. 1689, Springer, Berlin, 367-374.
- [20] Chima, Y.C., Kassima, A.A., Ibrahimb Y., 1999. Character recognition using statistical moments, *Image and Vision Computing* 17, 299-307.
- [21] 崔锦泰著, 程正兴译. 小波分析导论. 陕西: 西安交通大学出版社, 1997.
- [22] 丁险峰, 吴洪, 张宏江, 马颂德. 形状匹配综述. *自动化学报*, 2001, 27(5): 678-694.
- [23] Faloutsos C., Barber R., Flickner M., et al, Efficient and effective querying by image content. *Intelligent information systems*, 3(1): 231-262, 1994.
- [24] Farahnaz M., Mokhtarian F., 2001. Performance evaluation of corner detection algorithms under similarity

- and affine transforms. In: Proceedings of British Machine Vision Conference, Manchester, UK, 353-362.
- [25] Freeman H., 1961. On the encoding of arbitrary geometric configurations. IRE Trans. on Electronic Computers. 260-268.
- [26] Freeman H., 1974. Computer processing of line drawing images. ACM Comput. Surveys, 6: 57-97.
- [27] Freeman H., Davis L.S., 1977. A corner finding algorithm for chain coded curves. IEEE Trans. Computers, 26(3): 297-303.
- [28] Freeman H., 1978. Shape description via the use of critical points. Pattern Recognition, 10: 159-168.
- [29] Freeman H., Comparative analysis of line drawing modeling schemes. Computer Graphics Image Process, 12: 203-223, 1980.
- [30] Fudos I., Palios L., An efficient shape-based approach to image retrieval. PRL, 23(6): 731-741, 2002.
- [31] Granlund G., 1972. Fourier preprocessing for hand print character recognition, IEEE Trans. Comput. 21: 195-201.
- [32] Gupta, L., Srinath, M.D., 1987. Contour sequence moments for the classification of closed planar shapes, Pattern Recognition, 20(3): 267-272.
- [33] He X.C., Yung N.H., 2004. Curvature scale space corner detector with adaptive threshold and dynamic region of support. In: Proceedings of the 17th IEEE International conference on Pattern recognition, 791-794.
- [34] Hermilo S.-C., Ernesto B., Ramon R.D., 2007. Efficiency of chain codes to represent binary objects. Pattern Recognition 40, 1660-1674.
- [35] 洪安祥. 基于内容的图像检索若干论题研究. 博士学位论文, 浙江大学, 2003.
- [36] Hu M.K., 1962. Visual pattern recognition by moment invariants. IRE Trans. on Information Theory, 8(2): 179-187.
- [37] Huang J., 1998. Color-Spatial Image Indexing and Applications. Phd thesis, Cornell University, New York, USA.
- [38] 黄元元. 基于视觉特征的图像检索技术研究. 博士论文, 南京理工大学, 2003.
- [39] Iivarinen J., Visa A., 1996. Shape recognition of irregular objects, In: Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, SPIE, 25-32.
- [40] Jeannin S. (Ed.), 2000. MPEG-7 Visual part of experimentation model version 5.0, ISO/IEC JTC1/SC29/WG11/N3321, Noordwijkerhout, March, 2000.
- [41] 金振玉. 信息论. 北京: 北京理工大学出版社, 1991.
- [42] Kauppinen H., Seppanen T., Pietikainen M., 1995. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification, IEEE Trans. Pattern Anal. Mach. Intell. 17(2) : 201-207.
- [43] Kim Y.-S., Kim W.-Y., 1998. Content-based trademark retrieval system using a visually salient feature, Image Vision Comput. 16: 931-939.
- [44] Kim W.-Y., Kim Y.-S., 2000. A region-based shape descriptor using Zernike moments, Signal Process. Image Commun. 16: 95-102.
- [45] Krzyzak A., Leung S.Y., Suen C.Y., 1989. Reconstruction of two dimensional patterns from Fourier Descriptors, Mach. Vision Appl. 2, 123-140.
- [46] Liao S. X., Pawlak M., 1996. Image analysis by moments, IEEE Trans. On Pattern Analysis and Machine Intelligence, 18(3): 254-266.
- [47] 林元烈. 应用随机过程. 北京: 清华大学出版社, 2002, 78-130.
- [48] 刘进, 张天序. 图像不变矩的推广. 计算机学报, 2004, 27(5): 668-674.
- [49] Liu, J., Zhang, T.X., 2004. Fast algorithm for generation of moment invariants, Pattern Recognition, 37: 1745-1756.
- [50] 刘淑娟. 可变夹角链码的研究. 硕士学位论文, 河北师范大学, 2005.

- [51] Liu Y.K., Zalik B., 2005. An efficient chain code with Human coding. *pattern Recognition* 38, 669-676.
- [52] Liu, Y.S., Yang L.H., Sun Q., 2004. Contour-based moment invariants and their application to the recognition of object shapes, *Journal of Image and Graphics*, 9(3): 308-313.
- [53] Loncric S., 1998. A survey of shape analysis techniques, *Pattern recognition*, 31(8): 983-1001.
- [54] Lu G. J., Sajjanhar A., 1999. Region-based Shape Representation and Similarity Measure Suitable for Content-based Image Retrieval. *Multimedia System*, 7(2): 165-174.
- [55] Manjunath B.S., Salembier P., Sikora T., 2002. Introduction to MPEG-7: Multimedia content description interface, John Wiley & Sons, Inc., New York, NY.
- [56] Meier T., 1998. Segmentation for video object plane extraction and reduction of coding artifacts, PhD Thesis, University of Western Australia, Australia.
- [57] Mokhtarian F., Mackworth A.K., 1992. A theory of multiscale: curvature based shape representation for planar curves. *IEEE Trans. Patern Analysis Machine Intelligence*, 14 (8): 789-805.
- [58] Mokhtarian F., Abbasi S., Kittler J., 1996a. Robust and efficient shape indexing through curvature scale space, *Proceedings of the British Machine Vision Conference*, Edinburgh, UK, 53-62.
- [59] Mokhtarian F., Abbasi S., Kittler J., 1996b. Efficient and robust retrieval by shape content through curvature scale space, *International Workshop on Image Databases and Multimedia Search*, Amsterdam, The Netherlands, 35-42.
- [60] Mokhtarian F., Suomela R., 1998. Robust image corner detection through curvature scale space. *IEEE Traps Pattern Analysis and Machine Intelligence*, 20(12): 1376-1378.
- [61] Mokhtarian F., Mohanna F., 2001. Enhancing the curvature scale space corner detector. *Proceedings of Scandinavian conference on Image Analysis*. Norway: Bergen, 145-152.
- [62] Ohm J.R., Bunjamin F.B., Liebsch W., et al, 2000. A set of visual feature descriptors and their combination in a low-level description scheme, *Signal Process. Image Commun.* 16: 157-179.
- [63] Pentland A, Picard R, Sclaroff S., 1994. Photobook: tools for content-based manipulation of image database. *Proc. SPIE*, 2185, 34-47.
- [64] Peura M., Iivarinen J., 1997. Efficiency of simple shape descriptors, in: *Proceedings of the Third International Workshop on Visual Form*, Capri, Italy, 443-451.
- [65] Prokop R. J., Reeves, A.P. 1992. A survey of moment-based techniques for unconcluded object representation and recognition. *Proceedings of international conference on automatic face and gesture recognition*, 3(2): 438-460.
- [66] Rattarangsi A., Roland T.C., 1983. Scale space filtering Karlsruhe: *Proceedings of the 8th international Joint Conference on Artrificial Intelligence*. 1019-1022.
- [67] Rattarangsi A., Roland T.C., 1992. Scale-based detection of comer of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (4): 430-449.
- [68] Ritendra D., Dhiraj J., Li J., et al., 2008. Image retrieval : ideas, influences, and trends of the new age. *ACM Transactions on Computing Survey*, 40(2): 1-66.
- [69] Rosenfeld A., Weszka J.S, 1975. An improved method of angel detection on digital curves. *IEEE Trans. Computers*. 1975, 24(9): 940-941.
- [70] Rucklidge W.J., 1997. Efficient locating objects using Hausdorff distance, *Int. J. Comput. Vision* 24 (3): 251-270.
- [71] Safar M., Shahabi C., Sun X., 2000. Image Retrieval by Shape: A Comparative Study. *IEEE Int. Conf. On Multimedia and Expo(I)*. USA, 141-144.
- [72] Sajjanhar A., 2003. Spatial information in histograms for shape representation, *Lecture Notes in Computer Science* 2690, 855-859.
- [73] Salem, A.B.M., Sewisy, A.A., Elyan, U.A., 2005. A Vertex Chain Code Approach for Image Recognition. *ICGST International Journal on Graphics, Vision and Image Processing*, 5(3): 1-8.

- [74] Sarkar D., 1993. A simple algorithm for detection of significant vertices for polygonal approximation of chain-coded curves, *pattern recognition letters*, 14(12): 959-964.
- [75] Scassellati B., Slexopoulos S., Flickner M., 1994. Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgments, in: *SPIE Proceedings on Storage and Retrieval for Image and Video Databases II*, Vol. 2185, San Jose, CA, USA, pp. 2-14.
- [76] 盛骤, 谢式千, 潘承毅. 概率论与数理统计. 北京: 高等教育出版社, 1989.
- [77] 孙君顶, 崔江涛, 毋小省等. 基于颜色和形状特征的彩色图像检索方法. *中国图像图形学报*, 2004a, 9(7): 820-827.
- [78] 孙君顶, 毋小省, 周利华. 基于信息熵的图像检索. *西安电子科技大学学报*, 2004b, 31(2): 223-228.
- [79] 孙君顶, 武学东, 周利华. 基于颜色和形状的图像检索. *计算机科学*, 2004c, 31(5): 180-183.
- [80] 孙君顶, 张喜民, 崔江涛等. 一种新的基于颜色和空间特征的图像检索方法, *计算机科学*, 2005, 32(6): 158-160, 184.
- [81] 孙君顶. 基于内容的图像检索技术研究. 博士学位论文, 西安电子科技大学, 2005.
- [82] 孙君顶, 毋小省. 基于分块主色和形状特征的彩色图像检索. *光电工程*, 2006, 33(12): 85-90.
- [83] Sun J.D., Wu X.S., 2006a. Chain code distribution-based image retrieval. In: *Intelligent Information Hiding and Multimedia Signal Processing*, 139-142.
- [84] Sun J.D., Zhang X.M., Cui J.T., et al., 2006b. Image Retrieval Based on Color Distribution Entropy. *Pattern Recognition Letters* 10, 1122-1126.
- [85] Sun J.D., 2006. Image retrieval based on improved entropy and moments. In: *Intelligent Information Hiding and Multimedia Signal Processing*, 347-350.
- [86] 孙君顶, 毋小省. 基于熵及不变矩特征的图像检索. *光电工程*, 2007, 34(6): 102-106, 115.
- [87] Sun J.D., Wu X.S., 2007a. Shape retrieval based on the relativity of chain codes. *International Workshop Proceedings, Multimedia Content Analysis and Mining (MCAM'2007)*, Weihai, China, 76-84.
- [88] Sun J.D., Wu X.S., 2007b., An Effective Approach towards Contour-Shape Retrieval, 2007 International Conference on Computational Intelligence and Security(CIS'2007), 2007, 335-338.
- [89] Sun J.D., Zhang Z.S., 2008. Shape retrieval based on combination moment invariants, *Proceedings of Information Technology and Environmental System Science*, Volume 3, 301-305.
- [90] 孙君顶. 基于链码分布特征及相关性的轮廓描述与检索. *光电子.激光*, 2008, 19(8): 1112-1115.
- [91] Teague M.R., 1980. Image analysis via the general theory of moments, *J. Opt. Soc. Am.* 70 (8): 920-930.
- [92] Teh C.-H., Chin R.T., 1988. On image analysis by the methods of moments, *IEEE Trans. Pattern Anal. Mach. Intell.* 10 (4): 496-513.
- [93] Tieng Q.M., Boles W.W., 1997. Recognition of 2D object contours using the wavelet transform zero-crossing representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 19(8): 910-916.
- [94] Veltkamp, R.C., Latecki, L.J., 2006. Properties and Performance of Shape Similarity Measures, In *Proceedings of IFCS 2006 Conference: Data Science and Classification*.
- [95] 王小玲, 谢康林. 一种新的方向码描述的图像检索方法. *哈尔滨工业大学学报*, 2006, 38(9): 1545-1548.
- [96] Wong, W.H., Siu, W.C., Lam K.M., 1995. Generation of moment invariants and their uses for character recognition, *Pattern Recognition Letters* 16 115-123.
- [97] Wu, C.H., Horng, S.J., Wen, C.F., Wang, Y.R., 2007. Fast and scalable computations of 2D image moments, *Image and Vision Computing*, 1-13.
- [98] Wu X.S., Sun J.D., 2008. Shape retrieval of irregular objects, 2008 *Proceedings of Information Technology and Environmental System Science*, Volume 3, 319-322.
- [99] 夏德深, 傅德胜. 现代图像处理技术与应用. 南京: 东南大学出版社, 1997.
- [100] 肖茜, 鲁宏伟. 基于高斯平滑的自适应角点检测. *计算机辅助设计与图形学学*, 2003, 15(11):

- 1358-1361.
- [101] 杨翔英, 章毓晋. 小波轮廓描述符及在图像查询中的应用. 计算机学报, 1999, 22(7): 752-757.
- [102] 姚玉荣, 章毓晋. 利用小波和矩进行基于形状的图像检索. 中国图像图形学报, 2000, 5A(3): 206-210.
- [103] Yong I., Walker J., Bowie J., 1974. An analysis technique for biological shape, Comput. Graphics Image Process. 25: 357-370.
- [104] Zhang D.S., 2002. Image Retrieval Based on Shape, PhD Thesis, Monash University.
- [105] Zhang D.S., Lu G., 2002a. Generic Fourier descriptor for shape-based image retrieval, in: Proceedings of IEEE International Conference on Multimedia and Expo. (ICME2002), Vol. 1, Lausanne, Switzerland, 425-428.
- [106] Zhang D.S., Lu G.J., 2002b. Shape-based image retrieval using generic fourier descriptor. Signal Processing: Image Communication, 17(10): 825-848.
- [107] Zhang D.S., Lu G.J., 2004. Review of shape representation and description techniques. Pattern Recognition 37, 1-19.
- [108] 张小洪, 雷明, 杨丹. 基于多尺度曲率乘积的鲁棒图像角点检测, 中国图像图形学报, 2007, 7(12): 1270-1275.
- [109] 章毓晋. 基于内容的视觉信息检索. 北京: 科学出版社, 2003.
- [110] 赵宇, 陈雁秋. 曲线描述的一种方法: 夹角链码. 软件学报, 2004, 15(4): 300-307.
- [111] 周明全, 耿国华, 韦娜. 基于内容图像检索技术. 北京: 清华大学出版社, 2007.
- [112] 钟宝江, 廖文和. 基于精化曲线累加弦长的角点检测技术. 计算机辅助设计与图形学学报, 2004, 16(7): 939-943.
- [113] 庄越挺. 智能多媒体信息分析与检索的研究. 博士学位论文, 浙江大学, 1998.
- [114] 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索. 北京: 清华大学出版社, 2002.
- [115] 章毓晋. 图像处理和分析 (上册). 北京: 清华大学出版社, 1999.

## 第5章 基于纹理特征的图像检索

纹理是一种不依赖于颜色和亮度的反映图像同质现象的视觉特征，体现了物体表面共有的内在特性；同时，纹理还包含了物体表面组织结构排列的重要信息以及它们与周围环境的联系。正因为如此，纹理特征在基于内容的图像检索中得到了广泛的应用。本章介绍了四类（统计法、结构法、频谱法和模型法）经典的纹理描述算法，在此基础上，详细讨论了基于纹理基元共生矩阵的描述方法。

### 5.1 图像的纹理描述

#### 5.1.1 纹理及纹理特征的定义

纹理是自然界中常见的现象，是任何事物构成成分的分布或特征。习惯上认为纹理是表现在视觉图像上灰度或颜色分布的某种规律性，这种规律性在不同类别的事物中有着不同的特点。人们常将那些在图像局部区域内呈现不规则性，而在整体上表现出某种规律性的特征称为纹理。它具有层次性、尺度性、平移不变性和确定与随机二重性。与其他图像特征相比，纹理反映了图像灰度模式的空间分布，包含了图像的表面信息及其与周围环境的联系，更好地兼顾了图像的宏观信息与微观结构，纹理分析便成为图像分析的重要手段。由于人们对纹理的感受是与心理效果相结合的，用语言或文字来描述纹理通常很困难，因此，尽管人们的视觉系统能很方便的识别纹理，但对纹理图像却很难下一个确切的定义。

根据霍金斯的纹理三大要素，Hawkins 给出了一个对纹理的详细描述（Hawkins J.K. 1970，阮秋琦 2001）：

- （1）某种局部的“样式”（序列性）在一个序列更大区域内的不断重复；
- （2）序列是由基本部分非随机排列组成的；
- （3）各部分大致都是均匀的统一体，纹理区域内任何地方都有大致相同的结构尺寸。

Coggins 收集了在计算机视觉领域中一些经典的纹理定义（Coggins J.M., 1985）：

- （1）纹理可以被认为是由肉眼可见的区域组成。纹理结构的简单特征是有重复图案的组成，在这些图案中的图元按一定的布局规则排列。
- （2）如果图像的一组局部统计特征或者其他特征是不变的，变化缓慢的或者近似周期的，那么就认为图像区域含有不变的纹理。

文献[Castleman K.R., 2002]也给出了一个对纹理特征比较好的解释：纹理是一种反映了图像中一块区域的像素灰度级的空间分布属性，这种空间结构的固有属性可以通过邻域像素之间的相关性进行刻画。

这些关于纹理的描述在感觉上是合理的，已经逐渐被广大学者所接受和应用。

根据纹理所采用描述方法的不同，还可以将纹理分为结构纹理（确定性过程）和统计纹理（随机过程）两种类型（王海霞 2006）分别进行定义。结构纹理是指纹理在某一确定



的图像区域中,以近乎周期性的种类和方式重复其自身的局部基本模式,在这一定义下,纹理由基本模式及其规则排列构成,这形成了结构法纹理分析的基础;统计纹理是指纹理在某一确定的图像区域中,相邻像素的灰度(或色调、颜色)服从某种统计排列形成的一种空间分布,这一定义要求通过统计方法实现对纹理的描述。

另外,纹理与尺度有密切的联系,要分析纹理,需要确定一定的尺度。一幅纹理图像在较粗的尺度下可能看不出纹理来,需要到更细的尺度上观察。而且,纹理具有区域性质的特点,通常被看做对局部区域中像素之间关系的一种度量,常使用区域的尺寸、可分辨灰度元素的数目以及这些灰度元素的相互关系来描述一幅图像中的纹理区域。因此,对单个像素来说,讨论纹理没有意义,纹理可用来辨识图像中的不同区域。从这个意义上来,纹理又可定义为在视场范围内的灰度分布模式。这是一种可操作的定义,可确定要分析纹理所需做的工作和所应采取的方法。

纹理特征,是指利用计算机技术从数字图像中计算出来的,可以定量描述人对纹理的定性的感知的某些参数,它对区域内部灰度变化或色彩变化的某种规律进行量化,这些纹理特征能够尽可能地缩小纹理的类内差距,同时尽可能增大纹理的类间差距。纹理的视觉特征一般有三个基本量:周期性、方向性和随机性,其中周期性和方向性是两个高层次的纹理特征,可以用来指导纹理图像的知觉感知。不同的应用问题和不同的图像类型都给图像纹理特征提出了不同的需求,一方面,纹理定义的不统一性使得纹理分析中的问题更为复杂,更具有挑战性;另一方面也由于纹理本身具有的多种属性使得研究者可以引入多种模型来描述纹理特征。

### 5.1.2 常用的纹理分析方法

纹理分析,是指通过一定的图像处理技术提取纹理特征,从而获得纹理的定量或定性描述的处理过程。一般包括两方面的内容,即检测出纹理基元和获得有关纹理基元的排列分布方式的信息。其基本过程是:先从像素出发,在纹理图像中提取出一些辨识能力比较强的特征,作为检测出的纹理基元,并找出纹理基元排列的信息,建立纹理基元模型;然后再利用这些纹理基元模型对纹理图像进行分割、分类或识别处理。目前已经有不少纹理特征提取算法,但是由于在实际处理的图像中纹理的随机性非常大,因此,很难找到一种广泛适用的纹理模型。近年来研究者们结合时频分析、分形学、神经网络等方法,对纹理分析提出了许多创新性较强的方法,很大程度上提高了纹理分析的精度。概括起来说,图像纹理特征的提取方法主要有四大类:统计法、结构法、模型法和频谱法。

#### (1) 统计法

统计法是利用像素间的局部相关性来刻画纹理,主要适用于分析像木纹、森林、山脉、草地那样的纹理细腻而且不规则的物体。该方法试图从纹理图像中计算出一些在某个区域内或区域间的边界处保持相对平稳的特征值,以此特征值作为特征来表示区域内的一致性以及区域间的相异性。多数统计方法都来源于 Julesz 的发现(Julesz B. 1975),他认为人类视觉系统是使用统计特征来区分不同纹理的,包括一阶统计量、二阶统计量以及更高阶统计量。这类方法的典型代表是灰度共生矩阵(Gray Level Co-occurrence Matrices, GLCM)的纹理特征分析方法(Haralick R.M. et al 1973);另一种典型方法则是通过对图像的能量谱函数计算来提取纹理的粗细度及方向性等特征参数(Brochard J. et al 2001),还有一种方法是直接从图像的自相关函数来提取纹理特征。

## (2) 结构法

结构法是建立在纹理基元理论基础上的—种纹理分析方法。纹理基元理论认为,复杂的纹理可以由若干简单的纹理基元以—定的有规律的形式重复排列构成。结构法利用基元排列的规则性来分析纹理,主要包含两个要素:基元特征和基元组合规则。有如下两种策略:①计算各个基元的统计特征作为纹理特征,如平均强度、面积、周长、方向、离心率等;②基于基元组合规则分析与合成纹理,有图模型法、树文法等。该方法的主要问题是纹理基元的提取。结构法只适用于纹理基元较大而有规则的物体,对自然纹理很难取得较满意的效果,因此结构分析法的应用受到了很大程度的限制。

## (3) 模型法

模型法假设纹理按某种模型分布,采用模型的参数作为纹理特征。典型的方法是随机场模型法,如马尔可夫随机场(MRF)模型法、Gibbs 随机场模型法、Wold 模型、分形模型法等。这类方法的关键是如何为待处理图像选择适合的模型以及如何估计所选模型的参数。该类方法存在计算量大,自然纹理很难用单一模型表达的缺点。

## (4) 频谱法

统计方法和基于模型的方法与人类视觉机理脱节,难以进行更精确的纹理描述,因此,在分析纹理图像时,常常采用频谱法。频谱法就是将纹理图像看做二维信号,使用滤波方法对纹理图像进行分析的—种方法。该方法将图像分解为不同的频率和方向成分,采用时频多分辨纹理分析法,利用多分辨率多通道特征来对图像作频率和方向选择性滤波,得到相应特征。显然,这类方法的成功与否取决于滤波算子对目标特征的提取效果,因此应该根据不同种类的特征设计更合理的滤波器。频谱法主要包括傅里叶变换法、Gabor 变换法和小波变换法等。

对纹理采用哪种表达和描述方法依赖于纹理的模式和尺度,根据纹理的不同模式或尺度需要采用不同的方法。如果模式中包含许多小的纹理基元,就可以对它进行统计分析而不考虑其纹理基元;如果模式中包含许多大的纹理基元,可以对它基于纹理基元进行结构分析;如果模式中包含许多小的纹理基元但它们组成—些小的聚类,则可以采用统计的方法进行图像分割将各个聚类检测出来,此时不用考虑纹理基元,在此基础上,利用检测出来的聚类对模式进行结构分析;如果模式中包含构成聚类的大的纹理基元,通过将纹理基元组合,可以用结构的方式将聚类检测出来,然后基于检测出来的聚类可对模式进行结构分析。

### 5.1.3 纹理的分类

由于目前对纹理没有—个统一的定义,因而根据纹理不同方面的属性,纹理的分类方式也有很多种。

人们通常认为纹理是由许多相互接近的、互相编织的元素构成(它们常常富有周期性),所以直观地说纹理描述可提供图像区域的平滑、稀疏、规则性等特性。而从心理学方面来讲,人类观察到的纹理特征包括粒度、方向性和重复性等。根据这些特性的不同,—般可将纹理分为自然纹理与人工纹理。自然纹理是在自然界中大量存在着的纹理,如草地、森林、沙漠等等,是具有重复现象的自然景象,其纹理是不规则的;人工纹理则是由自然背景上的符号排列组成的,如砖块,格子窗等,是有规则的。而从纹理的构成上则可以将纹理描述成确定的和随机的两类。确定性的纹理图案通常是—人造的,多由点、线、多边形等

有规律的排列组成,而随机纹理图像则通常是自然界产生的。

按纹理的模式和尺度,通常情况下可将纹理分为三类[Rao A.R. 1990]:

(1) 全局有序纹理 (Globally Ordered Texture), 也称强 (Strong) 纹理, 这种纹理或者包含对某些纹理基元的特定排列, 或者由同一类基元的特定分布构成。

(2) 局部有序纹理 (Locally Ordered Texture), 也称弱 (Weak) 纹理, 这种纹理在其中的每个点存在某种方向性。

(3) 无序纹理指那些既无重复性也无方向性的纹理。

全局有序纹理常可用结构分析法来分析。无序纹理有可能基于不平整度来描述, 用统计法分析比较合适。而局部有序纹理具有局部的方向性, 但在全图中方向是随机的, 或者说是各向异性的, 故采用统计法或结构法都不易建模, 具有一定的特殊性。这类纹理图像的方向场可看做包含两幅图像, 一幅是角度 (Angle) 图, 另一幅是相干 (Coherence) 图[Rao A. R. 1990, 章毓晋 2007]。角度图表达了纹理中各点的主要局部朝向, 而相干图则表达了纹理中各点的各向异性情况。

#### (1) 角度图

方向场可看做一幅本征图(章毓晋 2007)。对纹理局部方向的估计主要有如下 5 个步骤。

① 用高斯滤波器平滑图像 (预处理)。

② 计算平滑后图像的梯度。

③ 利用切线的倒数计算局部朝向角度, 这步和边缘检测不同: 在边缘检测中, 梯度角是有反正切函数计算的, 但如果用反正切函数计算朝向角则无法确定唯一的角; 利用切线的倒数来计算局部朝向角度可将结果限制在  $[-\pi/2, \pi/2]$  范围, 且对纹理朝向的表达是唯一的。

④ 在给定邻域中求局部朝向角的平均值。

⑤ 计算模式相干 (即局部朝向估计的相近程度) 的测度。这里测量相干的基本思路是将某个邻域的方向矢量投影到该邻域中某个代表性的方向矢量上并归一化结果。如果朝向是相干的, 那么归一化投影将接近 1, 如果朝向是不相干的, 那么投影将互相抵消, 归一化结果将更接近于零。

#### (2) 相干图

令  $\theta(x, y)$  代表在  $(x, y)$  点估计出的朝向角,  $G(x, y)$  代表在  $(x, y)$  点的梯度幅度, 如图 5.1。

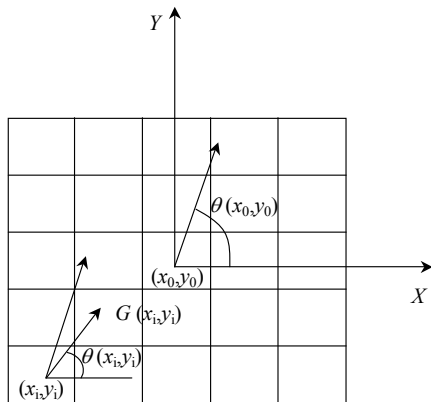


图 5.1 计算纹理方向场的示意图 (章毓晋 2007)

为确定在点  $(x_0, y_0)$  的相干值, 考虑处在围绕点  $(x_0, y_0)$  的给定尺寸的窗口  $W$  中的点  $(x_i, y_i)$ 。将沿  $\theta(x_i, y_i)$  方向的梯度幅度  $G(x_i, y_i)$  投影到沿  $\theta(x_0, y_0)$  方向的单位矢量上, 为  $G(x_i, y_i) \cos[\theta(x_0, y_0) - \theta(x_i, y_i)]$ 。对所有  $W$  中的  $(i, j)$  值, 计算投影绝对值 (利用绝对值以避免相反方向的矢量相抵消) 的和, 这个和给出对相干性的一个测度。如果对各点  $(x_i, y_i)$  的投影绝对值再用在该点的梯度幅度  $G(x_i, y_i)$  加权 (在有高视觉反差, 即大梯度的地方应该有高的相干性), 就能得到一个更好的对相干性的测度, 即

$$\rho = G(x_0, y_0) \frac{\sum_{(i,j) \in W} |G(x_i, y_i) \cos[\theta(x_0, y_0) - \theta(x_i, y_i)]|}{\sum_{(i,j) \in W} G(x_i, y_i)} \quad (5-1)$$

#### 5.1.4 纹理研究及应用领域

自然界的纹理无处不在, 人造物体表面更是规则纹理的体现, 利用纹理的表现和区分能力, 可以更好地服务科研、生活等各个方面。目前, 有关纹理方面的研究可分为以下 4 类 (章毓晋 2007)。

(1) 纹理表达和描述。对纹理特点进行刻画, 以表示纹理数据, 辨认纹理模式。

(2) 纹理分割。利用纹理作为特征对图像进行分割, 将图像分成具有相同或相似纹理的多个区域。

(3) 纹理分类与合成。根据对纹理的分解和分类 (对每个均匀的纹理区域赋予一个唯一的标号, 从而将各个区域划归为预定的类别), 利用对纹理的描述构建感知上与实际接近的纹理 (如将其覆盖在合成图形的表面, 可使图形产生真实感)。

(4) 由纹理恢复形状。根据图像的纹理特点 (以及纹理变化) 来恢复原始成像物体表面的朝向或表面的形状, 利用纹理信息提取表面朝向或深度的信息。

由于纹理是物体表面固有的一种特性, 图像区域常具有纹理性质, 所以纹理分析技术的应用范围也非常广泛, 主要有以下几个方面 (章毓晋 2007):

(1) 遥感遥测图像分析。遥感图像包括: 卫星图像、卫星多谱段图像、地震测量、声呐图像等, 图像中陆地、河流、海洋、农田、城市和山脉都具有纹理。主要应用包括: 遥感图像目标识别、遥感图像地形分类、遥感图像云类识别、海底地形匹配、地震模式识别等。

(2) 工业品的缺陷检测及质量控制。产品外表面的质量和纹理特征息息相关, 纹理分析在其中的应用包括纺织企业的织物疵点检测、织物缺陷检测、林业生产中原木内部缺陷检测和钢管缺陷检测等; 纹理分析的方法还能对工业产品的外观质量作出自动评定和检测, 其中包括纺织品外观等级评定、机械产品表面粗糙度非接触检测、电站锅炉水冷壁污染程度检测、烟叶等级检测等。另外, 在半导体器件的检测中, 可用表面纹理来确定所制造器件的完整性, 在岩相学和金相学中, 可用纹理特征来辨认样品。

(3) 研究金属形变。当金属发生形变时, 它的晶格会重新排列, 有些模式产生的概率会比较大, 这种模式的变化会导致纹理的产生和变化。许多材料的特性都与晶格的模式朝向有关, 所以根据纹理的模式可判定材料的特性。

(4) 医学图像分析。由于病症常常体现在图像纹理的变化之上, 因此纹理属性对根据医学图像进行诊断非常重要。利用纹理能成功检测出乳腺图像的病变, 使得乳腺癌的早期治疗成为可能; 通过老鼠活细胞切片的电子显微镜图像的纹理可以来识别癌症细胞等。

(5) 军用产品的设计。战争一般都在野外环境进行,因此为避免人员的装备轻易暴露,需要对它们进行伪装,应用纹理分析的知识设计以假乱真的伪装器材。其中包括人员、车辆、飞行器的迷彩设计、喷涂、火炮、舰艇、车辆的伪装网设计等。

(6) 文本分割。在一定距离之外,印刷文档的文字区域可以看成纹理区域,对于特定的语言,字体文本表现为特定的纹理,使用纹理分析方法从扫描文本的图像区域分割出文本区域,成为印刷文档自动数字化的重要步骤。

(7) 基于内容的图像检索。随着因特网和数字技术的飞快发展,人们所接触的图像的数目和种类越来越多,为了有效地浏览和使用这些图像,需要使用功能强大的图像检索工具,而纹理则是进行图像检索的一个非常重要的特征依据,可以根据图像间纹理特征的相似程度,从数据库中找出在视觉上最接近的图像。

### 5.1.5 纹理描述存在的问题

尽管纹理是计算机视觉和图像处理领域中一个的重要研究内容,而且研究者已经做了大量的工作,提出了许多有价值的研究成果,但目前仍然存在许多挑战性的问题等待解决。

(1) 由于纹理具有广泛性和多样性,因而对纹理还没有一个大家都能接受的精确定义。如第一节所述,虽然有众多学者给出了一些纹理的定义,但这些不同的定义依赖于不同学者对纹理的不同的理解和具体的应用,并未获得大家一致的认可。

(2) 纹理分析方法种类繁多,但至今还没有一个特别有效的方法在纹理分类方面取得令人满意的性能。许多已发表的文献上的纹理分类方法,在做实验验证时大都选用标准纹理测试集 Brodatz 的一个子集,测试子集的选用对各种方法的评估影响至关重要,在 Brodatz 全集上的性能测试相对于在子集上,尤其是一个小的子集上的测试更加令人信服。然而,许多方法在 Brodatz 全集上的性能并不能令人满意,很少的文献在测试时选用 Brodatz 纹理全集作为测试数据。目前仍有许多学者在探索新的理论和方法,试图找到一个非常有效的纹理描述方法。

(3) 尽管现实中的场景都是彩色的,但目前有关纹理的研究大部分都集中在灰度纹理领域,彩色纹理的研究很少有人关注。目前,对彩色纹理分析还处于探索阶段,还没有一个非常有效的彩色纹理描述方法。

(4) 纹理分析方法尽管在理论探索方面取得了很多的成果,然而,这些方法由于理论性较高,不容易被其他领域的研究人员以及工程技术人员所理解和接受,使得许多方法并不能成功地应用到其他领域。

随着纹理分析探索的进一步深入,纹理分析的系统理论框架将变得愈加清晰,有关纹理研究领域存在的一系列难题将会得到很好的解决。

## 5.2 统计法纹理分析

统计法纹理分析就是根据人们的直观视觉感受,从心理因素出发,形成纹理特征的一种表示方法。其主要思想是,通过图像中灰度级分布的随机属性来描述纹理特征,比较适合描述自然纹理,常可提供纹理的平滑、稀疏、规则等性质。主要用于分析木纹、沙地、草坪等纹理细且不规则的图像。

### 5.2.1 直方图的矩

最简单的描述纹理统计特征的方法就是借助于灰度直方图的矩来描述纹理 (Lowitz G.E. 1983)。假设用  $H(r)$  表示一个灰度直方图, 则其  $n$  阶矩可表示如下。

$$\mu_n(r) = \sum_{i=1}^L (r_i - m)^n H(r_i) \quad (5-2)$$

式中,  $L$  表示直方图的维数,  $m$  表示直方图的均值,  $\mu_n$  与  $H(r)$  的形状有直接关系, 但这些矩与纹理在空间的绝对位置无关。在直方图的  $n$  阶矩中, 方差  $\mu_2$  是灰度对比度的度量, 表达了曲线相对于均值的分布情况, 描述了直方图的相对平滑程度, 体现了图像中灰度的分散程度;  $\mu_3$  可定义偏度, 它表达了曲线相对于均值的对称性, 描述了直方图的偏斜程度, 即直方图分布对称与否的情况;  $\mu_4$  可定义峰度, 表示直方图的相对平坦性, 即直方图分布聚集于均值附近还是接近两端的情况, 描述了图像中纹理灰度的反差。更高阶矩的物理意义不直接, 但也可以定量地描述直方图的特点。直方图矩的识别能力比较低, 如图 5.2 所示, 这两种不同的纹理图像具有相同的直方图, 即具有相同的直方图矩, 因此仅借助于直方图的矩特征就无法进行很好的区分。

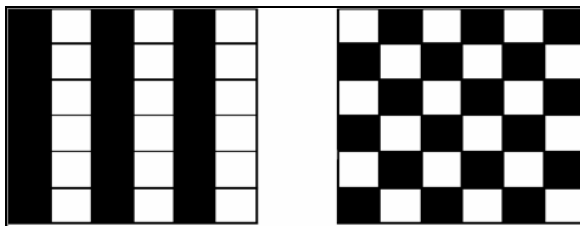


图 5.2 相同直方图特征的两幅纹理图像

### 5.2.2 二维灰度直方图

描述纹理特征的另一种简单方法是二维灰度直方图方法。二维直方图体现了部分位置信息, 用像素周围邻域的灰度差值作为当前像素灰度级。定义相邻像素的距离或位移向量为

$$d = (\Delta x, \Delta y) \quad (5-3)$$

式中,  $\Delta x$  和  $\Delta y$  是整数。相距为  $d$  的像素灰度差为

$$Y_d = |I(i, j) - I(i + \Delta x, j + \Delta y)| \quad (5-4)$$

对于不同的  $d$  可得到一组  $Y_d$ , 对给定的  $d$  统计  $Y_d$  得到差值直方图  $H(y_d, d)$ , 该直方图体现了一定的图像纹理信息。如果图像中的纹理较粗, 距离  $d$  与纹理基元尺寸相比较小, 直方图将聚集在低端; 如果纹理较细, 距离  $d$  与纹理基元相近, 则直方图的分布将会扩展开来。

从二维直方图  $H(y_d, d)$  提取的纹理属性特征有:

(1) 对比度

$$A_1 = \sum_{y_d=y_1}^{y_l} y_d^2 H(y_d, d) \quad (5-5)$$

(2) 角的二阶矩

$$A_2 = \sum_{y_d=y_1}^{y_t} [H(y_d^2, d)] \quad (5-6)$$

(3) 熵

$$A_3 = - \sum_{y_d=y_1}^{y_t} H(y_d, d) \log H(y_d, d) \quad (5-7)$$

(4) 均值

$$A_4 = \left( \frac{1}{t} \right) \sum_{y_d=y_1}^{y_t} y_d H(y_d, d) \quad (5-8)$$

式中,  $y_1 \leq y_d \leq y_t$ 。

### 5.2.3 灰度共生矩阵

由于纹理是相邻像素或相邻小区域灰度上及几何位置等相互关系的表征, 因此处于同样位置关系的一对像素的某种条件概率就可以用来表示其纹理特征。Haralick (Haralick R.M. et al 1973) 等人从数学角度研究了图像纹理中灰度级的空间依赖关系, 根据图像中各像素之间的角度方位和距离关系构造了一个灰度共生矩阵 (Gray Level Co-occurrence Matrix, GLCM)。该矩阵按照图像灰度值的空间关系描述像素点对之间的空间结构特征及其相关性, 表示相距  $(\Delta x, \Delta y)$  的两个灰度像素同时出现的联合分布概率。如图 5.3 所示, 其中,  $\Delta x$  和  $\Delta y$  的范围由像素间距  $\delta$  和方向  $\theta$  两个参数决定, 即  $\Delta x = \delta \cos \theta, \Delta y = \delta \sin \theta$ 。该方法可以表示纹理的稀疏度、对比度、复杂度及其纹理力度等特性。

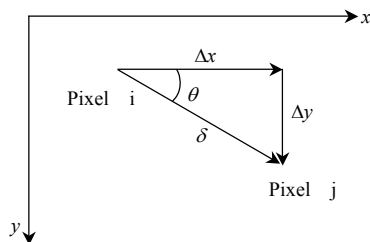


图 5.3 像素对分布示意图

设  $f(x, y)$  为一幅灰度图像, 对图像中的任一区域  $R$ , 定义  $S$  为区域中具有特定空间联系的像素对的集合, 则灰度共生矩阵可表示为

$$m_{(d, \theta)}(i, j) = \text{card} \{ [(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = i \& f(x_2, y_2) = j \} \quad (5-9)$$

式中,  $x_2 = x_1 + d \cos \theta, y_2 = y_1 + d \sin \theta$ ,  $\text{card}(S)$  表示集合  $S$  中对  $m_{(d, \theta)}(i, j)$  有贡献的元素个数。在实际应用中, 常需要对式 (5-9) 进行归一化, 即

$$m_{(d, \theta)}(i, j)' = \frac{\text{card} \{ [(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = i, f(x_2, y_2) = j \}}{\text{card}(S)} \quad (5-10)$$

为减少计算量, 计算中往往需要先对图像进行灰度变换, 降低灰度级, 同时减少  $\theta$  的方向数, 通常取  $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$  四个方向。

假设图像的灰度级有 8 级, 则对图 5.4 (a) 中所示的图像区域, 其共生矩阵  $m_{(1, 90^\circ)}$  如

图 5.4 (b) 所示。

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	0	0	0	0	0
2	0	1	0	2	0	0	0	0
3	0	0	1	1	0	0	0	0
4	0	1	0	0	1	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

图 5.4 灰度共生矩阵示例

共生矩阵体现了不同纹理之间的区别，对于具有不同特点的图像纹理，其共生矩阵明显不同。灰度共生矩阵中的主对角线上元素是一定位置关系下的两像素相同灰度值组合出现的次数。由于存在沿纹理方向上相近像素的灰度值基本相同，垂直纹理方向上相近像素间灰度值差异较大的一般规律，因此，这些主对角线元素的大小有助于判别纹理的方向和粗细，对纹理分析起着重要的作用。对纹理较为粗糙的图像，像素一般具有相同的灰度，其灰度共生矩阵中的值主要集中于主对角线附近；而对于纹理较为细腻的图像，由于其像素对灰度差异较大，其灰度共生矩阵中的值则散布在各处。例如，共生矩阵  $m_{(1,0)}(i, j)$  描述一幅图像水平方向上的灰度级的联合分布，如果这个共生矩阵的主对角线上的元素全部为0，则说明水平方向没有相邻的两个像素具有相同的值，即水平方向上灰度变化频繁，纹理较为细腻；如果主对角线上元素很大，则表明水平方向上纹理较为粗糙。

灰度共生矩阵把像素作为纹理基元，表示基元的特征是像素位置和灰度，基元之间的空间关系也被局限于空间位移向量，虽然在一定程度上能够很好地体现图像中的纹理信息，但在某些情况下对图像的特征描述常常需要对灰度共生矩阵进行推广。纹理基元可选用图像中的低层单元，如边缘点、边缘段，或灰度均匀的小区域等。描述这些纹理基元的特征可以是多个，如边缘单元可用位置、对比度、方向、灰度均值、灰度方差、区域尺寸和形状等信息来描述。同样，纹理基元之间的空间关系可以推广为用任意复杂的空间约束来定义的通用关系（章毓晋 2007）。

设  $U$  为图像中所有基元的集合,  $V$  为图像中基元性质的集合,  $f$  为赋予  $U$  中基元一个  $V$  中性质的函数。在基元间利用距离或相邻性建立空间联系, 设  $S \subseteq U \times U$  为满足上述空间联系的基元对间的联系集合, 则推广的共生矩阵  $M$  的元素定义为

$$m(v_1, v_2) = \frac{card\{ (u_1, u_2) \in S \mid f(u_1) = v_1 \ \& \ f(u_2) = v_2 \}}{card(S)} \quad (5-11)$$

$m(v_1, v_2)$  为满足上述空间联系的基元对 (一个基元具有性质  $v_1$ , 而另一个基元具有性质  $v_2$ )



的相对频率。 $\text{card}(S)$  的定义同式 (5-9)。

灰度共生矩阵是分析图像局部模式和排列规则的基础, 为了有效利用灰度共生矩阵所提供的图像灰度方向、间隔和变化幅度的信息, 文献 (Haralick R.M. et al 1973, Russ 2002) 在共生矩阵的基础上提取了一些有意义的统计量, 基于这些数字统计量, 对图像的纹理特征进行了定量描述。所提取的纹理特征主要有以下几种。

(1) 角二阶矩 (能量)。该特征反映了图像区域的均匀性或平滑性。

$$W_1 = \sum_i \sum_j \{m(i, j)\}^2 \quad (5-12)$$

在均匀区域, 灰度值变化较小, 大部分像素对具有相同或相近的取值, 主要发生在矩阵的对角线附近, 其他大部分元素为零; 而在非均匀区, 灰度值变化大的像素对较多, 在整个灰度共生矩阵上概率均匀分布, 而且矩阵中元素的值都很小, 所以, 非均匀区的角二阶矩比均匀区域的角二阶矩要小。该测度对区域内部有无灰度值变化较敏感, 但对灰度值变化数值大小不敏感, 即具有高的局部灰度值对比度的区域角二阶矩值不一定高。当共生矩阵中所有  $m(i, j)$  都相等时,  $W_1$  达到最小值。

(2) 对比度。又称非相似性, 可理解为图像的清晰度, 即纹理清晰度。

$$W_2 = \sum_i \sum_j (i - j)^2 m(i, j) \quad (5-13)$$

$|i - j|$  表示图像特定位置关系下像素对的灰度值差, 灰度值差大的像素对越多, 这个值就越大。对比度反映了近邻像素的反差, 当图像中两个灰度级点对的统计个数接近共生矩阵对角线时, 纹理变化较小, 对比度较小; 反之, 则表明近邻像素的反差较大, 纹理较细。因此对比度值的大小反映了纹理的粗细度。

(3) 相关性。相关系数在一定程度上反映了矩阵行与列的线性相关程度。相关系数较大时图像区域灰度分布比较均匀。

$$W_3 = \frac{\sum_i \sum_j ij m(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (5-14)$$

式中,  $\mu_x, \mu_y, \sigma_x, \sigma_y$  分别定义为:

$$\begin{aligned} \mu_x &= \sum_i i \sum_j m(i, j) \\ \mu_y &= \sum_j j \sum_i m(i, j) \\ \sigma_x^2 &= \sum_i (i - \mu_x)^2 \sum_j m(i, j) \\ \sigma_y^2 &= \sum_j (j - \mu_y)^2 \sum_i m(i, j) \end{aligned} \quad (5-15)$$

(4) 熵。给出了图像内容随机性的度量, 是图像具有信息量的度量。

$$W_4 = - \sum_i \sum_j m(i, j) \log_2 [m(i, j)] \quad (5-16)$$

纹理信息是图像信息的一种, 若图像没有任何纹理, 则灰度共生矩阵几乎为零矩阵, 熵值接近于 0; 若图像有较多的细纹理, 则矩阵中元素近似相等, 该图像的熵值最大; 若仅有较少的纹理, 则矩阵中的元素的差别较大, 图像的熵值就较小。

(5) 差分矩。定义如下:

$$W_5 = \sum_i \sum_j (i - \mu)^2 m(i, j) \quad (5-17)$$

式中,  $\mu$  为  $m(i, j)$  的均值。

(6) 逆差分矩。又称均匀性。

$$W_6 = \sum_i \sum_j \frac{m(i, j)}{1 + (i - j)^2} \quad (5-18)$$

对于匀质区域, 其灰度共生矩阵的元素集中在对角线上,  $(i - j)^2$  值小, 则均匀性特征值较大; 对非匀质区域, 由于其灰度共生矩阵的元素集中在远离对角线上,  $(i - j)^2$  值大, 则均匀性的值较小, 所以该特征是图像分布平滑性的测度。

(7) 和平均。

$$W_7 = \sum_k \sum_i \sum_j im(i, j) \quad (5-19)$$

式中,  $k = i + j$

(8) 和方差。

$$W_8 = \sum_k \sum_i \sum_j (i - W_7)^2 m(i, j) \quad (5-20)$$

式中,  $k = i + j$

(9) 和熵。

$$W_9 = - \sum_k \sum_i \sum_j m(i, j) \log_2 m(i, j) \quad (5-21)$$

(10) 差方差。

$$W_{10} = \sum_{k=2}^{2N} \sum_{i=1}^N \sum_{j=1}^N m(i, j), \quad k = |i - j| = 0, 1, \dots, N - 1 \quad (5-22)$$

式中,  $d = \sum_{i=1}^N \sum_{j=1}^N |i - j| m(i, j)$ 。

由于上述分量的物理意义和取值范围不同, 因此检索时需要对它们进行内部归一化。这样, 在进行相似度度量时, 可使各分量具有相同的权重。高斯归一化是一种较好的归一化方法, 其特点是少量超大或超小的元素值对整个归一化后的元素值的分布影响较小 (Ortega M. 1997)。

在共生矩阵基础上定义的纹理描述符能够对图像特征进行很好的描述, 但其最大的缺陷在于这些统计特征纯粹从数学角度推算而来, 和人们在视觉上对纹理特征的鉴别和感知无法建立对应关系; 同时, 由于灰度共生矩阵本身具有方向性, 从矩阵中提取的统计量也只能反应某一方向信息, 虽然可以采用几个方向的统计量做平均, 但还是无法很好表达图像的空间信息, 对于包含不同纹理区域的图像, 无法体现图像中各纹理区域之间的空间位置关系; 而且, 从矩阵中提取的特征矢量还不具备尺度不变性。

为了解决利用共生矩阵进行检索时存在的上述问题, 文献 (Takahashi N. et al 2000) 通过将彩色图像转换成灰度图像进行分块和压缩, 提取各个分块的共生矩阵表示纹理特征, 然后对这些共生矩阵计算表示对比度和相关性等属性的 7 个统计量进行图像相似比较。 (Gagaudakis G. et al 2002) 等人构造了具有旋转不变性的循环共生矩阵 (Circular Co-Occurrence Matrix), 用于产生图像的纹理直方图, 来辅助进行基于颜色及其空间分布的检索。文献 (杨

育彬 2003) 则从人类的视觉感知入手判断颜色的相似性, 进而对图像进行分割, 构造图像连通区域, 提出了基于颜色连通区域的颜色共生矩阵用于描述图像的纹理特征。文献(王波等 2006) 考虑了原始图像灰度级与平滑图像灰度级的联合统计分布, 提出了广义图像灰度共生矩阵概念, 把原图像与平滑图像相对应的两个像素点合起来, 形成一个二元矢量, 把所得的二元矢量的集合称为广义图像, 然后提取图像的纹理特征。

#### 5.2.4 灰度-梯度共生矩阵

描述图像的特征除了利用灰度本身之外, 还有一个重要的方法是灰度变化的梯度。图像灰度大小构成了图像的基础, 而图像梯度则是构成图像轮廓、边缘的要素。文献(洪继光 1984) 结合图像的灰度和梯度特征, 提出了灰度-梯度共生矩阵方法。

为了避免过高的灰度级给分析带来庞大计算量, 首先对灰度和梯度作正规化处理, 使其最高灰度级为  $N_g$  级。对原始图像  $\{f(i, j), i=1, 2, \dots, N_x, j=1, 2, \dots, N_y\}$  作如下正规化变换:

$$F(i, j) = \left[ f(i, j) N_g / f_{\max} \right] + 1 \quad (5-23)$$

式中,  $[x]$  表示  $x$  的整数部分,  $f_{\max}$  为图像的最高灰度值。变换后的图像  $F(i, j)$  的灰度分布在离散间隔更大的等分  $N_g$  个级中。

设  $g(i, j)$  为图像梯度矩阵(采用 Sobel 算子进行轮廓提取), 对它作正规化处理, 使得梯度值分布在更大的离散间隔  $N_s$  个等级中, 即

$$G(i, j) = \left[ g(i, j) N_s / g_{\max} \right] + 1 \quad (5-24)$$

式中,  $g_{\max}$  为图像的最大梯度值。

经灰度-梯度正规化变换后, 正规化灰度矩阵取值范围是  $\{1, 2, \dots, N_g\}$ ; 正规化梯度矩阵取值范围是  $\{1, 2, \dots, N_s\}$ 。从而, 灰度-梯度共生矩阵可以定义如下:

$$\{H(i, j); i=1, 2, \dots, N_g, j=1, 2, \dots, N_s\} \quad (5-25)$$

式中,  $i, j$  表示矩阵的第  $i$  行第  $j$  列元素,  $H(i, j)$  表示正规化后图像中同时具有第  $i$  级灰度第  $j$  级梯度的总像素点数。

对灰度-梯度矩阵作正规化处理, 得到

$$P(i, j) = H(i, j) / (N_x N_y) \quad (5-26)$$

类似于灰度共生矩阵算法, 采用灰度-梯度共生矩阵进行图像纹理特征描述时也需要从中提取统计量。针对灰度-梯度共生矩阵, 提取的统计量有以下几种。

(1) 小梯度优势

$$T_1 = \left[ \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{p(i, j)}{j^2} \right] / H \quad (5-27)$$

(2) 大梯度优势

$$T_2 = \left[ \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} j^2 p(i, j) \right] / H \quad (5-28)$$

(3) 灰度分布的不均匀性

$$T_3 = \sum_{i=1}^{N_g} \left[ \sum_{j=1}^{N_s} p(i, j) \right]^2 / H \quad (5-29)$$

(4) 梯度分布的不均匀性

$$T_4 = \sum_{j=1}^{N_s} \left[ \sum_{i=1}^{N_g} p(i, j) \right]^2 / H \quad (5-30)$$

在这四个式子中,  $H = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i, j)$ 。

(5) 能量

$$T_5 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} [p(i, j)]^2 \quad (5-31)$$

(6) 相关

$$T_6 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} (i - \sigma_F)(j - \sigma_G)p(i, j) \quad (5-32)$$

式中

$$\begin{aligned} \sigma_F &= \left\{ \sum_{i=1}^{N_g} (i - \mu_F)^2 \left[ \sum_{j=1}^{N_s} p(i, j) \right] \right\}^{1/2} \\ \sigma_G &= \left\{ \sum_{j=1}^{N_s} (j - \mu_G)^2 \left[ \sum_{i=1}^{N_g} p(i, j) \right] \right\}^{1/2} \\ \mu_F &= \sum_{i=1}^{N_g} i \left[ \sum_{j=1}^{N_s} p(i, j) \right] \\ \mu_G &= \sum_{j=1}^{N_s} j \left[ \sum_{i=1}^{N_g} p(i, j) \right] \end{aligned} \quad (5-33)$$

(7) 灰度熵

$$T_7 = - \left\{ \sum_{i=1}^{N_g} \left[ \sum_{j=1}^{N_s} p(i, j) \right] \log_2 \left[ \sum_{j=1}^{N_s} p(i, j) \right] \right\} \quad (5-34)$$

(8) 梯度熵

$$T_8 = - \left\{ \sum_{j=1}^{N_s} \left[ \sum_{i=1}^{N_g} p(i, j) \right] \log_2 \left[ \sum_{i=1}^{N_g} p(i, j) \right] \right\} \quad (5-35)$$

(9) 惯性

$$T_9 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} (i - j)^2 p(i, j) \quad (5-36)$$

### 5.2.5 纹理谱

数字图像的每一个像素（除边界像素外）都由 8 个相邻像素包围着，这个  $3 \times 3$  邻域是表示其中心像素所有 8 个方向的最小完整单元。纹理谱方法是通过描述中心像素与其周围邻域像素之间的相对灰阶关系而提出的，其主要思想是：采用局部纹理模式作为纹理单元来进行纹理分析，通过刻画图像像素点邻域内灰度的变化来描述图像的纹理结构特征。因此，单个像素的纹理可从其周围的  $3 \times 3$  邻域中提取。与以往方法相比，纹理谱方法方便简单，所需的计算量较少，近年来越来越受到重视。

一个给定的 $3 \times 3$ 邻域,如图5.5所示,含有9个元素,记为 $V = \{V_0, V_1, \dots, V_8\}$ ,  $V_0$ 表示中心像素的灰度值,  $V_i$ 是邻域中第 $i$ 个像素的灰度值。定义一个相应的纹理单元 $T$ ,  $T$ 是一个包含8个元素的组合集(He Dong-chen et al 1990, Karkanis S. et al 1999, 盛文等 2000),

$$T = \{E_1, E_2, \dots, E_8\} \quad (5-37)$$

其中,  $E_i$ 定义为

$$E_i = \begin{cases} 0, & V_i \leq (V_0 - \Delta) \\ 1, & (V_0 - \Delta) < V_i \leq (V_0 + \Delta) \\ 2, & V_i > (V_0 + \Delta) \end{cases} \quad (5-38)$$

式中,  $i=1,2,3,\dots,8$ ,  $\Delta$ 表示一个小的正常量,元素 $E_i$ 占据像素 $i$ 相同的位置。由于 $T$ 中每个元素具有3种可能的取值,所有8个元素的总和共有 $3^8 = 6561$ 个可能的纹理单元,纹理单元数的定义如下:

$$N = \sum_{i=1}^8 E_i \times 3^{i-1} \quad (5-39)$$

这些纹理单元描述了一个给定像素的“本”纹理特征,即中心像素与其周围邻域像素之间的相对灰阶关系。对图像中所有纹理单元的发生频数进行统计,这个发生频数函数就表征了所要分析的图像的纹理信息,称这个所有纹理单元的频数发生函数为纹理谱。用横坐标表示纹理单元函数,纵坐标表示频数就可形成纹理谱直方图。一幅图像通常由两部分组成,纹理基元和随机噪声或背景。纹理成分相对于背景的百分比越大,纹理特征越容易被人的视觉所察觉。就纹理谱而言,一幅图像的纹理成分百分比越大,就会导致特殊的峰分布,而且,不同的纹理是由具有不同纹理谱的特殊纹理单元所组成。从这一点来看,一幅图像的纹理信息可以由纹理谱来描述。

对于图5.5所示的 $3 \times 3$ 像素邻域还有一种更为简单的纹理谱模型。该模型采用了不同的纹理单元,每个元素只有两种取值可能,即0或1,所以 $N$ 共有 $2^8 = 256$ 个可能取值。该模型降低了时间和空间上的复杂度,能更好地应用于多媒体图像检索中。

$V_1$	$V_2$	$V_3$
$V_4$	$V_0$	$V_5$
$V_6$	$V_7$	$V_8$

图5.5 图像中的一个 $3 \times 3$ 像素邻域

### 5.2.6 LBP算法

在上述纹理谱方法的基础上,(Ojala T. et al 1996, Ojala T. et al 2002)提出了LBP(Local Binary Pattern)方法,该算法在纹理分类上取得了很好的效果,近年来,广泛应用于医学图像处理及人脸识别等领域。

LBP算法通过刻画图像中每个像素点与其邻域内其他各点的灰度值的差异来描述图像纹理的局部结构特征,该局部结构可以用一个二进制的数字来量化。这种以邻域为单位的

某种局部结构可以看做是一个纹理单元，该纹理单元在整幅图像中有规律的出现就构成了一定的纹理，而对整幅图像中纹理单元的统计则可以表示整幅图像的纹理特征。从某种意义上说，该方法将局部的纹理结构信息以及全局的纹理统计信息同时融合到纹理分析中，为同时分析图像中随机的微观纹理和确定的宏观纹理提供了一个有效的工具。

LBP 算法的基本原理是对于一个  $3 \times 3$  的窗口，将中心像素与其邻域像素进行比较，若周围像素值大于中心像素值，则将该点赋值为 1，否则赋值为 0，最后将一个权值模板与阈值处理后的图像进行对应相乘求和，得到中心像素的值。如图 5.6 所示，则中心像素的 LBP 值可表示如下

$$\text{LBP} = 1 + 16 + 32 + 64 + 128 = 241 \quad (5-40)$$

例图	阈值	权值																																					
<table><tr><td>6</td><td>5</td><td>2</td></tr><tr><td>7</td><td>6</td><td>1</td></tr><tr><td>9</td><td>8</td><td>7</td></tr></table>	6	5	2	7	6	1	9	8	7	<table><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td></td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	1	0	0	1		0	1	1	1	<table><tr><td>1</td><td>2</td><td>4</td></tr><tr><td>128</td><td></td><td>8</td></tr><tr><td>64</td><td>32</td><td>16</td></tr></table>	1	2	4	128		8	64	32	16	<table><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>128</td><td></td><td>0</td></tr><tr><td>64</td><td>32</td><td>16</td></tr></table>	1	0	0	128		0	64	32	16
6	5	2																																					
7	6	1																																					
9	8	7																																					
1	0	0																																					
1		0																																					
1	1	1																																					
1	2	4																																					
128		8																																					
64	32	16																																					
1	0	0																																					
128		0																																					
64	32	16																																					

图 5.6 LBP 算法计算示意图

为了描述不同尺度下的纹理结构，可使用圆形邻域  $(P, R)$  ( $P$  表示圆形邻域中像素个数， $R$  表示邻域半径) 以及对像素的线性插值，将算子扩展到任意的半径以及任意邻域像素个数。

设  $C$  是位于坐标  $(x, y)$  处的一个像素，以  $C$  为中心以  $R$  为半径的环上等间隔分布的  $P$  个点构成了一个中心对称的环形局部结构，称之为  $P$  邻域， $P$  的数目可以根据间隔的大小来确定。当  $P=4, R=1$  时，该结构包含与像素点  $C$  水平和垂直相邻的 4 个像素，对应于数字图像处理中常用的 4-邻域，当  $P=8, R=1$  时，像素  $C$  的 4 个相邻的对角像素与其 4-邻域组成了该像素点的 8-邻域。可以看出一个局部的  $P$  邻域中共包含  $P+1$  个像素。图 5.7 分别表示了圆形  $(8, 2)$  邻域和圆形的  $(8, 3)$  邻域。

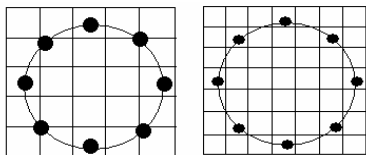


图 5.7 圆形的  $(8, 2)$  邻域和  $(8, 3)$  邻域

由于上述算法对纹理特征的描述非常有限，Ojala 又提出了对于原始的 LBP 算法的另一个推广—uniform 模式，就是当把二进制串看做一个圆时，串中从 0 到 1 以及从 1 到 0 的转换不超过 2。具体定义如下：

设单色调图像局部邻域的纹理特征为  $T$ ， $P(P > 1)$  个图像像素的灰度联合分布为

$$T = t(g_c, g_0, \dots, g_i, \dots, g_{P-1}) \quad (5-41)$$

式中，灰度值  $g_c$  是局部邻域的中心值， $g_i$  是在半径为  $R(R > 0)$  的圆形对称区域内第  $i$  个相邻空间像素灰度值。如果  $g_c$  的坐标是  $(0, 0)$ ，那么  $g_i$  的坐标是  $(-R \sin 2\pi/P, R \cos(2\pi/P))$ ，当邻域值不能精确地落到像素的中心，采用双线性二插值的方法进行估计。

由于用相邻像素与中心像素的差值代替相邻像素的像素值,可以实现旋转不变性,于是,将对称的环形邻域中的值  $g_i (i = 0, 1, \dots, P-1)$  减去中心像素的灰度值  $g_c$ , 式 (5-41) 变为

$$T = t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_i - g_c, \dots, g_{P-1} - g_c) \quad (5-42)$$

该纹理结构不仅包含了邻域中的各个像素自身的灰度值信息,还包含了中心像素和其邻域内其他像素点之间的差值信息,是中心像素灰度值和差值信息的联合分布。

假设差值  $g_i - g_c$  是独立于  $g_c$  的,则式 (5-42) 的分布可以表示为

$$T = t(g_c) t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (5-43)$$

在实际中,由于  $g_i - g_c$  和  $g_c$  是不可能绝对独立的,所以,式 (5-43) 中的分布形式只能是式 (5-42) 表示的联合分布的一种近似。但是为了得到有用的信息可以允许少量的信息丢失。式 (5-43) 中的  $t(g_c)$  描述了整个图像中所有图像灰度值的分布,对于局部纹理分析来说,它的意义不是很大,因此,式 (5-42) 表示的图像的局部邻域的灰度联合分布中包含的纹理属性的大部分信息可以用差值  $g_i - g_c$  的联合分布来近似代替 (Ojala et al 2002),即

$$T \approx t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (5-44)$$

差值  $g_i - g_c$  会受灰度尺度的影响,即当灰度值成比例变化时,得到的差值是不同的,这种差值的变化使得  $g_i - g_c$  整体的分布不具有稳定性。为了消除这种尺度变化的影响,可以通过只考虑差值的符号而不是它的实际值来得到一个尺度不变性的算子,即

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)) \quad (5-45)$$

式中,

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (5-46)$$

此时得到的分布中所有元素的值都为 0 或 1,这个局部结构可以看做一个纹理单元。uniform 模式 LBP 的测度则采用 0 和 1 空间变换的数目来实现,定义为

$$LBP_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_s) & \text{if } U(LBP_{P,R}) \leq 2 \\ p+1 & \text{otherwise} \end{cases} \quad (5-47)$$

$$U(LBP_{P,R}) = |s(g_{p+1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (5-48)$$

其计算过程如图 5.8 所示 (周明全等 2007)。

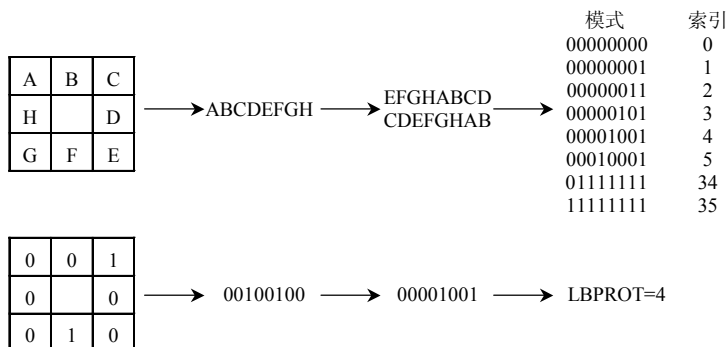


图 5.8 uniform 模式 LBP 的计算过程 (周明全等 2007)

Ojala 的实验表明,使用尺度为(8,1)的 LBP 算子时,uniform 模式的数量为 59,占 LBP 算子总数的 23%,却可以表达出其中 87.2%的纹理。可见这种模式包含了图像局部纹理的信息,可以有效地描述图像的大部分纹理特征,并明显减少特征的数量。

### 5.2.7 Laws纹理能量

纹理能量方法是由 Laws 在 1980 年提出的一种纹理滤波分析方法 (Laws K.L. 1980),它利用模板 (也称核) 计算局部纹理能量来获得灰度变化的信息。该方法首先对图像进行直方图均衡化处理,在  $3 \times 3$  或  $5 \times 5$  窗口内用模板进行微观滤波计算,取卷积之和作为输出新图像,然后将卷积后的图像中每个像素值用该像素为中心的  $15 \times 15$  局部窗口内的均方差或绝对值平均代替。Laws 称窗口内的均方值度量 “纹理能量方法”。定义如下:

设图像为  $f(x, y)$ , 一组模板为  $M_1, M_2, \dots, M_N$ , 则卷积  $g_n = f * M_n, n=1, 2, \dots, N$  给出了各个像素邻域中表达纹理特性的纹理能量分量。如果采用尺寸为  $k \times k$  的模板,则对应第  $n$  个模板的纹理图像 (的元素) 为

$$T_n(x, y) = \frac{1}{k \times k} \sum_{i=-(k-1)/2}^{(k-1)/2} \sum_{j=-(k-1)/2}^{(k-1)/2} |g_n(x+i, y+j)| \quad (5-49)$$

这样对应每个像素位置  $(x, y)$ , 都有一个纹理特征矢量  $[T_1(x, y), T_2(x, y) \dots T_N(x, y)]^T$ 。

设  $L$  代表层 (level),  $E$  代表边缘 (edge),  $S$  代表形状 (shape),  $W$  代表波 (wave),  $R$  代表纹 (ripple),  $O$  代表震荡 (oscillation), 则对应于常见的  $3 \times 3, 5 \times 5, 7 \times 7$  模板的 1-D 矢量 (写成行矢量) 形式分别如式 (5-50) 所示, 其中,  $L_5$  给出中心加权的局部平均,  $E_5$  检测边缘,  $S_5$  检测点,  $R_5$  检测波纹。

$$\begin{aligned} L_3 &= [1 \quad 2 \quad 1] \\ E_3 &= [-1 \quad 0 \quad 1] \\ S_3 &= [-1 \quad 2 \quad -1] \\ L_5 &= [1 \quad 4 \quad 6 \quad 4 \quad 1] \\ E_5 &= [-1 \quad -2 \quad 0 \quad 2 \quad 1] \\ S_5 &= [-1 \quad 0 \quad 2 \quad 0 \quad -1] \\ W_5 &= [-1 \quad 2 \quad 0 \quad -2 \quad 1] \\ R_5 &= [1 \quad -4 \quad 6 \quad -4 \quad 1] \\ L_7 &= [1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1] \\ E_7 &= [-1 \quad -4 \quad -5 \quad 0 \quad 5 \quad 4 \quad 1] \\ S_7 &= [-1 \quad -2 \quad 1 \quad 4 \quad 1 \quad -2 \quad -1] \\ W_7 &= [-1 \quad 0 \quad 3 \quad 0 \quad -3 \quad 0 \quad 1] \\ R_7 &= [1 \quad -2 \quad -1 \quad 4 \quad -1 \quad -2 \quad 1] \\ O_7 &= [-1 \quad 6 \quad -15 \quad 20 \quad -15 \quad 6 \quad -1] \end{aligned} \quad (5-50)$$

2-D 的模板可由计算 1-D 模板的外积得到。例如,



$$E_5^T L_5 = \begin{bmatrix} -1 \\ -2 \\ 0 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (5-51)$$

在许多实际应用中,常使用9个 $5 \times 5$ 的模板来计算纹理能量。当使用4个1-D矢量时可得到16个 $5 \times 5$ 的2-D模板,将这16个模板用于原始图像可得到16个滤波图像。设 $F_n(i, j)$ 为用第 $n$ 个模板在 $(i, j)$ 位置滤波得到的结果,那么对应第 $n$ 个模板的纹理能量图 $E_n$ 为( $c$ 和 $r$ 分别代表行和列)

$$E_n(r, c) = \sum_{i=r-5}^{r+5} \sum_{j=c-5}^{c+5} |F_n(i, j)| \quad (5-52)$$

每幅纹理能量图都是完全尺寸的图像,代表用第 $n$ 模板得到的结果。得到16幅纹理能量图后,有些对称的图像对可进一步结合(将一对图用它们的均值图代替),例如, $E_5 L_5$ 测量水平边缘而 $L_5 E_5$ 测量垂直边缘,它们的平均值将测量总的边缘。这样可得到9幅纹理能量图对应: $L_5 E_5 / E_5 L_5$ 、 $L_5 S_5 / S_5 L_5$ 、 $L_5 R_5 / R_5 L_5$ 、 $E_5 E_5$ 、 $E_5 S_5 / S_5 E_5$ 、 $E_5 R_5 / R_5 E_5$ 、 $S_5 S_5$ 、 $S_5 R_5 / R_5 S_5$ 、 $R_5 R_5$ ,这9幅纹理能量图也可看做一幅图,这样在其中每个像素位置都有一个含9个纹理属性的矢量。

Laws将Brodatz的8种纹理图像拼在一起,对图像作纹理能量测量,将每个像元指定为八个可能类中的一个,正确率达87%。这种纹理分析方法简单、有效,但所提供的模板较少,目前还没有更多地给出其变化性质,因此,应用受到一定的限制。

## 5.2.8 数学形态学分析法

数学形态学(Mathematical Morphology)是由随机集和整体几何理论的发展而形成的,目前已成为图像纹理分析的一个重要方法。

在数学形态学中的几个重要操作是腐蚀运算、膨胀运算以及由他们组成的开运算和闭运算。

(1) 用结构元素 $b$ 对输入图像 $f$ 进行灰度膨胀记为 $f \oplus b$ ,其定义为

$$(f \oplus b)(s, t) = \max \{ f(s-x, t-y) + b(x, y) \mid (s-t), (t-y) \in D_f \text{ 和 } (x, y) \in D_b \} \quad (5-53)$$

式中, $D_f$ 和 $D_b$ 分别是 $f$ 和 $b$ 的定义域。灰度图像膨胀运算的结果使图像中比背景亮的部分得到扩张,而比背景暗的部分受到收缩。膨胀计算是在由结构元素确定的邻域中选取 $f+b$ 的最大值,所以对灰度图像的膨胀操作有两类效果:①如果结构元素的值都为正,则输出图像会比输入图像亮;②如果输入图像中暗细节的尺寸比结构元素小,则其视觉效果会被减弱,减弱的程度取决于这些暗细节周围的灰度值以及结构元素的形状和幅值。

(2) 用结构元素 $b$ 对输入图像 $f$ 进行灰度腐蚀记为 $f \ominus b$ ,其定义为

$$(f \ominus b)(s, t) = \min \{ f(s+x, t+y) - b(x, y) \mid (s+t), (t+y) \in D_f \text{ 和 } (x, y) \in D_b \} \quad (5-54)$$

式中, $D_f$ 和 $D_b$ 的定义同上。灰度图像腐蚀运算的结果使比背景暗的部分得到扩张,而比背景亮的部分受到收缩。腐蚀计算是在由结构元素确定的邻域中选取 $f-b$ 的最小值,所以

对灰度图像的腐蚀操作有两类效果：①如果结构元素的值都为正，则输出图像会比输入图像暗；②如果输入图像中亮细节的尺寸比结构元素小，则其视觉效果会被减弱，减弱的程度取决于这些亮细节周围的灰度值以及结构元素的形状和幅值。

(3) 用  $b$  (灰度) 开启  $f$  记为  $f \circ b$ ，定义为

$$f \circ b = (f \ominus b) \oplus b \quad (5-55)$$

实际应用中，常用开启操作消除与结构元素相比尺寸较小的亮细节，而保持图像整体灰度值和大的亮区域基本不受影响。利用第一步的腐蚀去除小的亮细节并同时减弱图像的亮度，第二步膨胀增加（基本恢复）图像亮度但又不重新引入前面去除的细节。

(4) 用  $b$  (灰度) 闭合  $f$  记为  $f \bullet b$ ，定义为

$$f \bullet b = (f \oplus b) \ominus b \quad (5-56)$$

实际中常用闭合操作消除与结构元素相比尺寸较小的暗细节，而保持图像整体灰度值和大的暗区域基本不受影响。具体来说，采用第一步的膨胀去除小的暗细节并同时增强图像亮度，采用第二步的腐蚀减弱（基本恢复）图像亮度但又不重新引入前面去除的细节。

对图像进行数学形态学的运算和操作，一般先将图像进行二值化，得到二值化的图像后，利用具有某种形状和分布规则的称之为结构元素的平面点集（如一对有向点的距离，一有向线段，一个小圆等），对图像的局部模式进行某种集合运算，并求出结果集合的某种测试。然后用结构元素的参数（如一对有向点的距离  $d$ ，一有向线段  $l$ ，一个小圆  $r$  等）与这种测度之间的依赖关系来描述图像的纹理特征。对于一幅给定的二值图像和一种结构运算来说，如果结构元素不同，得到的结果也不一样。当结构元素依赖某变量  $\lambda$  的相似形时（如一对有向点的距离），二值图像的几何参量（如面积）的测定则依赖于  $\lambda$ 。我们称这种相依关系为二值图像关于给定结构运算的几何结构函数。在纹理分析过程中，常常使用几何结构函数来描述图像的纹理特征。常用的几何结构函数主要有三类（黄祥林 2001）。

(1) 协方差函数：结构元素为一长度为  $d$ ，方向为  $\alpha$  的有向点，用面积  $S(\bullet)$  作为几何参量的测定，则协方差函数的定义为  $C(d, \alpha) = S(A \ominus B)$ 。

(2) 线性矩函数：结构元素为一长度为  $l$ ，方向为  $\alpha$  的有向线段，用面积  $S(\bullet)$  作为几何参量的测定，则线性矩函数的定义为  $M(l, \alpha) = S(A \ominus B)$ 。

(3) 结构圆函数：结构元素为一半径为  $r$  的圆，用面积  $S(\bullet)$  作为几何参量的测定，则结构圆函数的定义为  $H(r) = S(A \ominus B)$ 。

其中， $A$  是二值图像中取 1 的点集， $B$  为结构元素。

这三个函数是数学形态学在纹理分析中使用的最基本函数，在实际应用中一般用这些函数曲线的特性来描述图像的纹理特征。如求  $C(d, 0)$  曲线的局部极大值点与极大值来分析和分类纹理图像。

### 5.2.9 自相关函数法

纹理结构常用其粗糙度来描述，其粗糙度与局部结构的空间重复周期有关。周期大的纹理粗，周期小的纹理称为细纹理。空间自相关函数是检验纹理粗糙测度的一种基本方法，纹理粗细度与自相关函数的关系如图 5.9 所示。纹理测度变化的倾向是小数值的纹理测度表示细纹理，反之表示粗纹理。

对于图像  $\{f(x_i, y_j); i, j = 0, 1, 2, \dots, N-1\}$ ，其自相关函数定义为

$$\rho(x, y) = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_i) f(x_i + x, y_i + y)}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_i)^2} \quad (5-57)$$

显然,  $0 \leq \rho(x, y) \leq 1$ , 当  $x=0, y=0$  时,  $\rho(x, y)=1$ , 取最大值, 利用自相关函数  $\rho(x, y)$  随  $x, y$  大小而变化的规律, 可以描述图像纹理的粗糙度、规整度以及粗略度等特征。

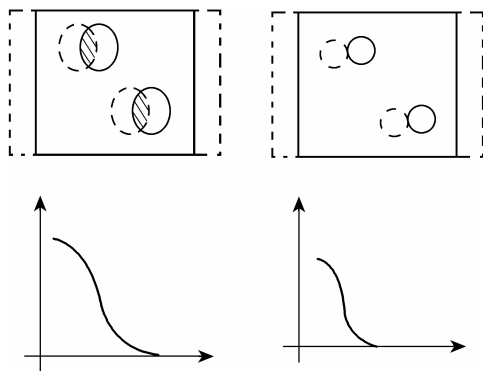


图 5.9 纹理粗细度与自相关函数的关系示意图

设

$$d = (x^2 + y^2)^{1/2} \quad (5-58)$$

若图像的纹理较粗, 则  $\rho(x, y)$  随  $d$  的增加而下降速度较慢; 若图像的纹理较细, 则  $\rho(x, y)$  随  $d$  的增加而下降速度较快。并且随着  $d$  的继续增加,  $\rho(x, y)$  会呈某种周期性变化, 其周期的大小可以作为描述图像纹理的重要特征, 能够反映局部模型排列规则的稀疏、稠密程度。对于具有比较规则的纹理图像, 自相关函数显示出峰值和谷值, 利用这种纹理特征进行分类有较好的效果。

### 5.2.10 Tamura纹理特征

基于人类对纹理的视觉感知研究, Tamura 等人提出了一种纹理特征的表达 (Tamura H. et al 1978)。该纹理特征和灰度共生矩阵的主要区别是 Tamura 纹理特征中所有纹理特性都在视觉上有意义, 而灰度共生矩阵的某些纹理属性不具有视觉意义 (如信息熵), 这一特点使得 Tamura 纹理特征在图像检索中使用较多。Tamura 纹理特征包括 6 个分量, 对应心理学角度上纹理特征的 6 种属性, 分别是粗糙度 (Coarseness)、对比度 (Contrast)、方向度 (Directionality)、线性度 (Linearity)、规整度 (Regularity) 和粗略度 (Roughness)。其中, 前 3 个分量对图像检索尤为重要, 下面着重讨论这三种特征的定义和数学表达。

#### (1) 粗糙度

粗糙度是反映纹理中粒度的一个量, 是最基本的纹理特征, 因此, 从狭义的观点来看, 纹理就是粗糙度。当两种纹理模式只是基元尺寸不同时, 具有较大基元尺寸的模式给人的感觉更粗糙。而对具有不同结构的纹理模式来说, 基元尺寸越大或者基元重复次数越少, 给人的感觉越粗糙。粗糙度可由不同大小窗口的像素的滑动均值得到, 计算可以分以下几个步骤进行。

首先, 计算图像中大小为  $2^k \times 2^k$  个像素的活动窗口中像素的平均强度值, 即

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i, j) / 2^{2k} \quad (5-59)$$

式中,  $k=0, 1, \dots, 5$ ;  $g(i, j)$  是位于  $(i, j)$  处像素的灰度值。

然后, 对于每个像素, 分别计算它在水平和垂直方向上互不重叠的窗口之间的平均强度差。

$$\begin{aligned} E_{k,h}(x, y) &= |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \\ E_{k,v}(x, y) &= |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})| \end{aligned} \quad (5-60)$$

取水平和垂直两方向上的最大均值差值为当前像素的邻域均值差值, 即

$$E_k(x, y) = \max(E_{k,h}(x, y), E_{k,v}(x, y)) \quad (5-61)$$

对于每个像素, 从多邻域尺寸中设置一个最佳尺寸, 即

$$S_{best}(x, y) = 2^k + 1 \quad (5-62)$$

使得  $E_k = E_{\max} = \max(E_1, E_2, \dots, E_L)$ , 其中  $E$  为邻域均值差值,  $L$  为邻域尺寸个数。

最后, 计算整幅图像中  $S_{best}$  的平均值作为灰度图像的纹理粗糙度, 表示为

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j) \quad (5-63)$$

式中,  $m$  和  $n$  分别是图像的宽度和高度。

显然, 整幅图像的粗糙度可以在一定程度上反映纹理图像的粒度特性, 但存在计算量大, 且衡量不准确等缺点。粗糙度特征的另一种改进形式是采用直方图来描述  $S_{best}$  的分布, 而不是如上述方法那样简单地计算  $S_{best}$  的平均值。这种改进后的粗糙度特征能够表达具有不同纹理特征的图像或区域, 对图像检索更为有利。

## (2) 对比度

对比度是反映像素亮度统计分布的一个量, 其大小由以下四个因素决定: 灰度级动态范围、直方图上黑白两部分两极化的程度、边缘的锐度和重复模式的周期。一般意义上, 我们说纹理的对比度是指前两个因素, 可以通过对像素强度分布情况的统计得到, 确切地说, 它是通过  $\alpha_4 = \mu_4 / \sigma^4$  来定义的, 其中  $\mu_4$  是四次矩而  $\sigma^2$  是方差。对比度通过如下公式衡量:

$$F_{con} = \frac{\sigma}{(\alpha_4)^n} \quad (5-64)$$

该值给出了整个图像或区域中对比度的全局变量,  $n$  通常取值为 8, 4, 2, 1, 1/2, 1/4 或 1/8。

## (3) 方向度

方向度指的是给定纹理区域的全局特性, 描述了纹理是如何沿某些方向散布或集中的。一般来说, 方向度与纹理基元的形状以及如何将这些纹理基元排列的规则有关, 它是基于像素的梯度向量计算的。计算方向度时, 首先计算每个像素处的梯度向量, 该向量的模和方向分别定义为

$$\begin{aligned} |\Delta G| &= (|\Delta_H| + |\Delta_V|) / 2 \\ \theta &= \tan^{-1}(\Delta_V / \Delta_H) + \pi / 2 \end{aligned} \quad (5-65)$$

式中,  $\Delta_H$  和  $\Delta_V$  分别是通过图像卷积下列两个  $3 \times 3$  操作符所得的水平和垂直方向上的变化量。

$$\begin{array}{ccc}
-1 & 0 & 1 \\
-1 & 0 & 1 \\
-1 & 0 & 1
\end{array}
\begin{array}{ccc}
1 & 1 & 1 \\
0 & 0 & 0 \\
-1 & -1 & -1
\end{array}
\quad (5-66)$$

Tamura 方法的核心就是构建方向角局部边缘概率直方图。当所有像素的梯度向量被计算出来后,对  $\theta$  值范围进行离散化,统计每个桶中相应的  $|\Delta G|$  大于给定域值的像素数量,构建方向角局部边缘概率直方图  $H$ 。该直方图对于具有明显方向性的图像会表现出峰值,对于无明显方向的图像则表现比较平坦。可以根据方向角局部边缘概率直方图计算直方图波峰尖锐程度来定量地描述纹理的方向性。Tamura 等人采用二阶矩累加的方法,其定义如下:

$$F_{dir} = 1 - r n_p \sum_p \sum_{\phi \in \omega_p} (\phi - \phi_p)^2 H(\phi) \quad (5-67)$$

式中,  $p$  代表直方图的峰值;  $n_p$  为直方图中波峰的个数;  $\omega_p$  代表峰值两侧谷底距离; 而  $\phi_p$  是波峰中心位置;  $r$  为直方图归一化系数;  $\phi$  为量化后的方向角;  $H$  为直方图。

### 5.2.11 灰度游程长度法

游程长度的概念源于数理统计,在数字图像中,定义沿一确定方向  $\theta$  具有相同灰度的一批邻接像素之集为一游程。简单地说,灰度游程长度即为同一直线上具有相同灰度值的像元集合,它与灰度级数、长度、方向等因素有关。一幅图像的游程长度矩阵反映了图像灰度关于方向、相邻间隔、变化幅度的综合信息,是分析图像的局部模式和像素点排列规则的基础。对于二维数字图像中任一像素点  $(x, y)$ ,一般计算如图 5.10 所示 8 个方向的游程。假设一幅  $5 \times 5$  的数字图像,如图 5.11 所示,其游程长度的计算结果如表 5-1 所示。

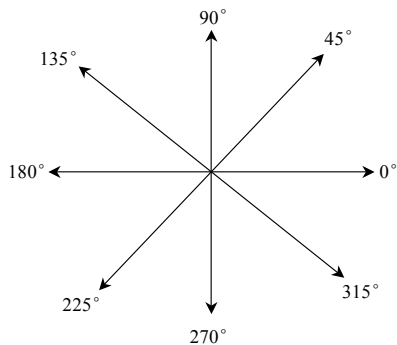


图 5.10 游程的 8 个方向

25	50	10	25	25
50	50	25	25	5
10	10	25	25	50
10	30	40	25	10
25	25	50	50	30

图 5.11  $5 \times 5$  的数字图像

表 5-1 游程长度计算结果

方向	0°	45°	90°	135°	180°	225°	270°	315°
游程长度	2	3	2	2	1	2	1	2

对于某一方向  $\theta$ , 其游程长度矩阵  $R(\theta)$  的元素定义为

$$r(i, j | \theta) = \# \{ (n_1, m_1), (n_2, m_2), \dots, (n_j, m_j) | g(n_k, m_k) = i \} \quad (5-68)$$

$\#(S)$  表示集合  $S$  中对  $r(i, j|\theta)$  有贡献的元素个数,  $\{(n_1, m_1), (n_2, m_2), \dots, (n_j, m_j)\}$  是沿方向  $\theta$ , 灰度值为  $i$ , 且长度为  $j$  的任一游程。显然,  $r(i, j|\theta)$  反映了图像中灰度值为  $i$ , 沿  $\theta$  方向, 游程为  $j$  的像素点个数。进而统计

$$r(i, j) = \sum_{\theta} r(i, j|\theta) \quad (5-69)$$

这样就构成了灰度游程长度矩阵  $R$ 。在求得灰度游程长度矩阵的基础上, Galloway (Galloway M. M. 1975) 定义了如下 5 个纹理特征。

(1) 短游程优势

$$RF_1(\theta) = \frac{1}{T_R} \sum_{i=0}^{N_G-1} \sum_{j=1}^{N_R} \frac{p(i, j/\theta)}{j^2} \quad (5-70)$$

(2) 长游程优势

$$RF_2(\theta) = \frac{1}{T_R} \sum_{i=0}^{N_G-1} \sum_{j=1}^{N_R} j^2 p(i, j/\theta) \quad (5-71)$$

(3) 灰度级非均匀性

$$RF_3(\theta) = \frac{1}{T_R} \sum_{j=1}^{N_R} \left[ \sum_{i=0}^{N_G-1} p(i, j/\theta) \right]^2 \quad (5-72)$$

(4) 游程长非均匀性

$$RF_4(\theta) = \frac{1}{T_R} \sum_{i=0}^{N_G-1} \left[ \sum_{j=1}^{N_R} j^2 p(i, j/\theta) \right]^2 \quad (5-73)$$

(5) 游程百分数

$$RF_5(\theta) = \frac{1}{T_R} \sum_{i=0}^{N_G-1} \sum_{j=1}^{N_R} p(i, j/\theta) \quad (5-74)$$

以上五个公式中,  $N_G$  为灰度级数,  $N_R$  为游程长个数,  $T_R$  为图像像素总数, 定义如下:

$$T_R = \sum_{i=0}^{N_G-1} \sum_{j=0}^{N_R} p(i, j/\theta) \quad (5-75)$$

## 5.3 结构法纹理分析

结构法纹理分析的基本思想是: 假定复杂的纹理模式是由简单的纹理基元(基本纹理元素)以一定的有规律的形式重复排列组合而成。基于这种假设, 1966年, Beck以不同的英文字母作为纹理基元进行观察(Beck J. 1966), 发现纹理基元按不同方向分布影响着人们对纹理的区分。1983年, Bergen和Julesz又在此基础上, 进行了一系列的精神物理实验(Bergen J.R. et al 1983), 发现不仅纹理基元的方向, 纹理基元的密度也显著影响着人们对不同纹理的区分。1988年, Bergen和Adelson又发现, 纹理基元的大小以及尺寸之间的对比, 也对纹理的区分有着重要的影响(Bergen J.R. et al 1999)。这一系列的实验, 从生理和心理的角度说明纹理图像可以分解为纹理基元, 从而奠定了以纹理基元的不同统计特征来分析纹理图像的基础。结构分析法适用于印刷图像, 如布料、砖墙类等人工形成的纹理图像, 其纹理元和排列比较规则, 按纹理基元的特性和其排列规则来描述。

### 5.3.1 结构法基本知识

在结构分析方法中,大体包括两方面的工作:其一,确定与提取基本的纹理单元,即纹理基元;其二,研究存在于纹理基元之间的“重复性”的结构关系。所以也有人认为纹理具有两层结构,第一层与表现灰度基元的局部性质有关,而第二层确定了灰度基元的组织情况。从这个意义上来说,特征提取就变为确定这些基元并定量地分析它们的排列规则。由于纹理基元描述了局部纹理特征,因此对整幅图像中不同纹理基元的分布统计可获得图像的全面纹理信息。

#### 1. 纹理基元

目前,没有标准的(或大家公认的)纹理基元集合。一般认为一个纹理基元是由一组属性所刻画的相连接的像素集合。最简单的基元就是像素,其属性就是其灰度。比它复杂一点的基元是一组均匀灰度相连的像素集合。选择纹理基元的原则就是简单规范。常用的纹理元素主要有椭圆和直线段,因此可以用椭圆的长、短轴长度,长轴方向,平均灰度等参数来作为纹理元素的特征。纹理基元可能是明确的、直观的,也可能是不明确的、需要人为根据情况设定的。例如,咖啡豆的纹理图像中,可以选用一粒粒咖啡豆作为纹理基元,也可以用各种尺寸的椭圆作为纹理基元。无论怎样确定纹理基元,为了提取它们(形成它们)都需要做图像的区域分割或边缘、线的提取。对于纹理基元之间的结构关系,可以有不同的分析途径。最简单的方法是分析纹理基元之间存在的相位、距离、尺寸等统计特性,也可以考虑用复杂的方法分析,诸如利用模型或形式语法等。

设纹理基元为  $h(x, y)$ , 排列规则为  $r(x, y)$ , 则纹理  $t(x, y)$  可表示为 (Jayaramamurthy 1980)

$$t(x, y) = h(x, y) \otimes r(x, y) \quad (5-76)$$

式中,  $r(x, y) = \sum \delta(x - x_m, y - y_m)$ 。

这里  $x_m$  和  $y_m$  是脉冲函数的位置坐标。根据卷积定理,在频域有

$$T(u, v) = H(u, v)R(u, v) \quad (5-77)$$

所以

$$R(u, v) = T(u, v)H(u, v)^{-1} \quad (5-78)$$

这样,给定对纹理基元  $h(x, y)$  的描述,可以推导反卷积滤波器  $H(u, v)^{-1}$ 。将这个滤波器用于待分割的纹理图像,得到在含有纹理区域中的脉冲阵列,每个脉冲都在纹理基元的中心。纹理基元描述了局部纹理特征,对整幅图像中不同纹理基元的分布进行统计可获得图像的全面纹理信息。这里可用纹理基元标号为横轴,以它们出现的频率为纵轴而得到纹理图像的直方图(纹理谱)(任仙怡等 1998, 章毓晋 2007)。

#### 2. 排列规则 (章毓晋 2007)

在获得纹理基元的基础上,还要建立将它们进行排列的规则。如果能定义出一些排列基元的规则,就有可能将给定的纹理基元按照规定的方式组织成所需的纹理模式。这里的规则和方式可用形式语法来定义。

假设有如下8个重写规则(其中,  $a$  表示模式,  $b$  表示向下,  $c$  表示向左):

- (1)  $S \rightarrow aA$  (变量  $S$  (起始符号) 可用  $aA$  来替换);
- (2)  $S \rightarrow bA$  (变量  $S$  (起始符号) 可用  $bA$  来替换);
- (3)  $S \rightarrow cA$  (变量  $S$  (起始符号) 可用  $cA$  来替换);

- (4)  $A \rightarrow aS$  (变量  $A$  可用  $aS$  来替换);
- (5)  $A \rightarrow bS$  (变量  $A$  可用  $bS$  来替换);
- (6)  $A \rightarrow cS$  (变量  $A$  可用  $cS$  来替换);
- (7)  $A \rightarrow c$  (变量  $A$  可用常量  $c$  来替换);
- (8)  $A \rightarrow a$  (变量  $A$  可用常量  $a$  来替换)。

结合使用不同的重写规则可生成不同的2-D模式。

设  $a$  是一个圆盘模式 (可看做一个纹理基元), 如图5.12 (a) 所示, 如果依次使用规则 (1), (4), (1), (4), (8) 可得到图5.12 (b), 如果依次使用规则 (1), (4), (1), (5), (3), (6), (3), (4), (1), (4), (2), (6), (3), (6), (1), (4), (8), 就得到图5.12 (c)。

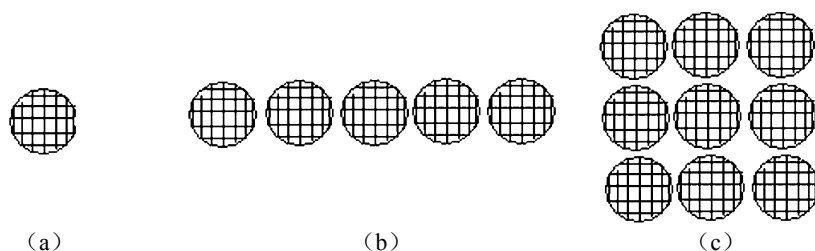


图 5.12 不同 2-D 模式的生成 (章毓晋 2007)

纹理区域的性质与基元的性质和数量都有关 (Haralick 1992)。如果一个小尺寸的图像区域包含灰度几乎不变的基元, 该区域的主要属性是灰度; 如果一个小尺寸的图像区域包含灰度变化很多的基元, 该区域的主要属性是纹理。这里的关键就是这个小图像区域的尺寸, 基元的种类以及各个不同的基元的数量和排列。当不同基元的数量减少时, 灰度特性将增强。事实上, 如果这个小图像区域就是单个像素, 则该区域只有灰度性质, 当小图像区域中不同基元的数量增加时, 纹理特征将增强; 当灰度的空间模式是随机的, 且不同基元的灰度变化比较大时, 会得到比较细的纹理; 当空间模式变得比较细, 且图像区域包含越来越多的像素时, 会得到比较粗的纹理。

### 5.3.2 纹理镶嵌

比较规则的纹理在空间中是以有次序的形式进行纹理单元的镶嵌, 最典型的模式是用 (一种) 正多边形镶嵌而成, 与采样网格的模式类似。图 5.13 给出了常见的几种模式。图 5.13 (a) 表示由正三角形构成的模式; 图 5.13 (b) 表示由正方形构成的模式; 图 5.13 (c) 表示由正六边形构成的模式。

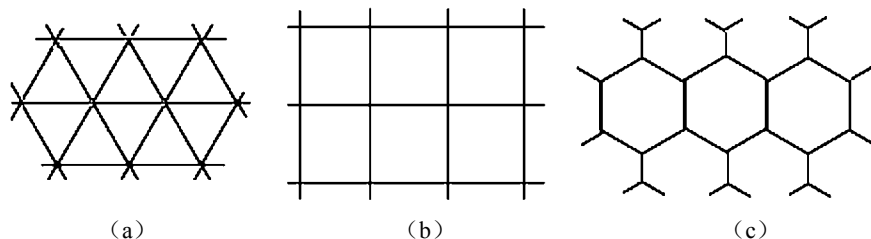


图 5.13 三种正多边形镶嵌 (章毓晋 2007)



如果同时使用两种边数不同的正多边形进行镶嵌就构成半规则镶嵌,图 5.14 给出了几种典型的半规则镶嵌模式 (Ballard D.H. et al 1982)。

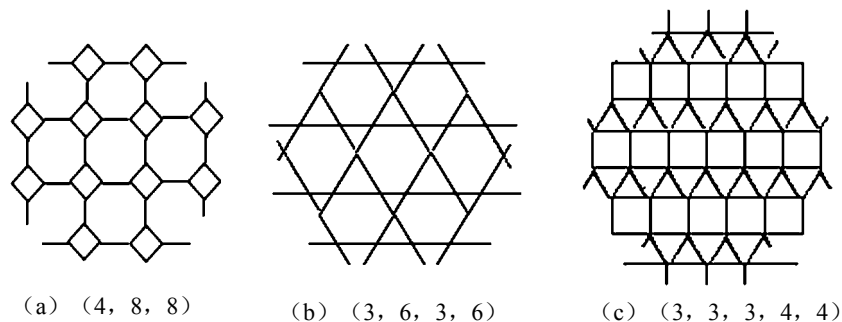


图 5.14 半规则镶嵌 (章毓晋 2007)

为描述上述镶嵌模式,可以依次列出绕顶点的多边形的边数。例如,对图 5.13 (c) 的模式,可表示为 (6, 6, 6),即对各个顶点都有 3 个六边形围绕它。对图 5.14 (c) 的模式,对各个顶点都有 3 个三角形和 2 个正方形围绕它,所以可以表示为 (3, 3, 3, 3, 4, 4)。这里重要的是基元的排列,而不是基元本身。对一个用基元来定义的镶嵌模式,排列所定义的镶嵌模式与它是对偶的。在图 5.15 中,图 5.15 (a) 和图 5.15 (c) 分别对应基元的镶嵌和排列的镶嵌,而图 5.15 (b) 是它们结合的结果 (章毓晋 2007)。

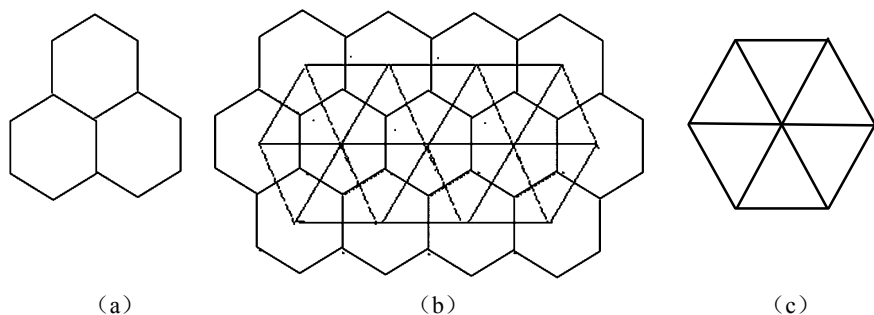


图 5.15 基元镶嵌和排列镶嵌的对偶性 (章毓晋 2007)

### 5.3.3 Voronoi多边形方法

利用 Voronoi 多边形也可以描述纹理 (Tuceryan M. et al 1990, Shapiro L. et al 2001)。假设已从图像中提取出一组纹理基元,对每个基元可用一个特殊点 (如重心) 来表达。如果令  $S$  是这些特殊点的集合,对  $S$  中任意一对点  $p$  和  $q$ ,可在它们之间画一条对分线。这条对分线将图像分成两半,其中一半包含与  $p$  比较近的点而另一半包含与  $q$  比较近的点。令  $H^q(p)$  代表与  $p$  比较近的那一半,如果对所有的  $q$  都如上进行,就可得到包含  $p$  的多边形,也称 Voronoi 多边形。相对于  $p$  的 Voronoi 多边形包含所有比  $S$  中任何点都近的点。可定义为

$$V(p) = \bigcap_{q \in S, q \neq p} H^q(p) \quad (5-79)$$

图 5.16 给出了相对于一组圆形纹理基元的 Voronoi 多边形。对处于图像内部的纹理基元得到的多边形为六边形,而对处于图像边缘的纹理基元得到的多边形具有不同的形状。

一旦从图像中提取出纹理基元并计算出 Voronoi 网格, 可以进一步计算形状特征从而将相同的多边形聚类, 例如, 图 5.16 中的纹理区域可用规则的六边形(形状特征)描述。

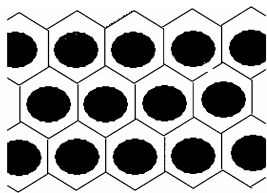


图 5.16 一组圆形纹理的 Voronoi 多边形(章毓晋 2007)

### 5.3.4 其他方法

纹理基元指的是数字图像中某像素与其周围邻域像素组成固定的基本结构形式(集合), 定义纹理基元的目的是试图从纹理结构方面来描述图像纹理特征。由于纹理基元可以由用户根据图像的内容重新定义, 近年来, 一些学者也提出了许多算法。

图 5.17 给出了文献(王亮申等 2004)所定义的几种纹理基元结构形式, 其中黑色方块为当前像素, 与其邻域像素(空心方块)构成不同的纹理基元。在此基础上, 对纹理基元的结构又进行了扩展, 如图 5.18 所示。这种方法把统计法和结构分析方法有机结合起来, 不仅仅描述了图像的灰度分布, 还结合了纹理基元的纹理结构描述, 起到了很好的效果。

文献(赵珊等 2006)借鉴方块编码的思想也构建了一种纹理基元模式, 如图 5.19 所示。在此基础上, 结合共生矩阵, 提出了纹理基元共生矩阵算法, 获得了较好的检索效果。

另外, Carlucci (Carlucci L. 1972)曾提出一个使用直线段、开放多边形和封闭多边形作为纹理基元的纹理模型, 其排列规则由一种图状语法结构定义。Lu (Lu S.Y. et al 1978)提出采用一种树型语法结构表示纹理, 他们将纹理按照  $9 \times 9$  的窗口进行分割, 每个分解单元的空间结构表示为一棵树。

总的来说, 由于实际的纹理大都是无规则的, 因此, 结构法在实际的应用中受到很大的限制。

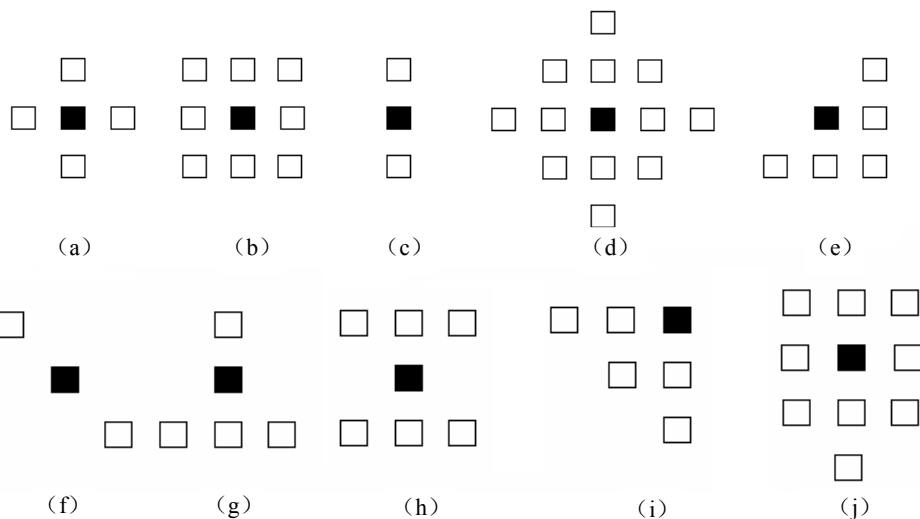


图 5.17 纹理基元结构图(王亮申等 2004)

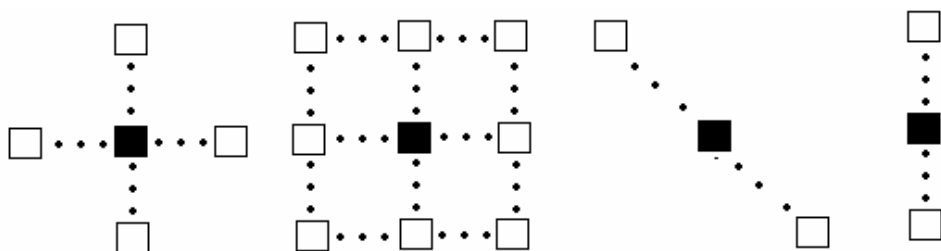


图 5.18 纹理基元结构的扩展 (王亮申等 2004)

图像块	<table><tr><td>20</td><td>22</td></tr><tr><td>8</td><td>7</td></tr></table>	20	22	8	7	<table><tr><td>9</td><td>19</td></tr><tr><td>11</td><td>20</td></tr></table>	9	19	11	20	<table><tr><td>17</td><td>11</td></tr><tr><td>18</td><td>9</td></tr></table>	17	11	18	9	<table><tr><td>20</td><td>7</td></tr><tr><td>6</td><td>23</td></tr></table>	20	7	6	23	<table><tr><td>8</td><td>7</td></tr><tr><td>6</td><td>8</td></tr></table>	8	7	6	8
20	22																								
8	7																								
9	19																								
11	20																								
17	11																								
18	9																								
20	7																								
6	23																								
8	7																								
6	8																								
二进制块	<table><tr><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td></tr></table>	1	1	0	0	<table><tr><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td></tr></table>	0	1	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td></tr></table>	1	0	1	0	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1
1	1																								
0	0																								
0	1																								
0	1																								
1	0																								
1	0																								
1	0																								
0	1																								
1	0																								
0	1																								
二进制码	1100	0101	1010	1001	1001																				
纹理值	12	5	10	9	9																				

图 5.19 图像块与相应的纹理基元值 (赵珊等 2006)

## 5.4 频谱法纹理分析

图像可以被视为在空间距离上有周期变化的信号。当周期固定时,其周期的倒数就表示为空间频率。在图像中,这个周期性变量表示了图像明暗变化的快慢,如图 5.20 所示。水平方向上三幅图都有明暗的变化。(a) 中空间频率低,(c) 中空间频率高,(b) 中的空间频率介于两者之间。

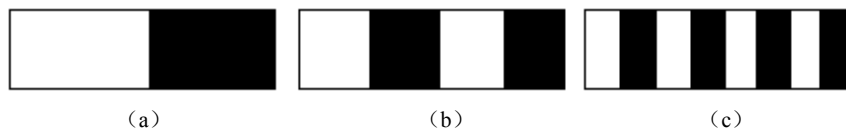


图 5.20 具有不同空间频率的图像

从图中我们可以看出,空间频率高的图像主要表征图像的细微变化或细节内容;空间频率低的图像则表征图像中大的物体轮廓或变化趋势。

频谱法主要借助于各种变换算法利用图像的频率特性来描述纹理特征。其关键主要是,寻求一种可逆的线性变换,从看似复杂的数据中找出一些直观的信息,再对它进行分析,从而可以用一组不相关的数据,通常是一组系数来代替图像数据。将这些系数按其含有图像信息以及对图像主观质量影响的重要程度排序,删除一些不会对图像内容描述产生重大影响的不重要系数,用少量的高效的系数进行图像的特征描述。由于图像信

号往往在频域具有比在空域更加简单和直观的特性，所以频谱分析法在图像分析中起着很重要的作用。

### 5.4.1 傅里叶变换法

傅里叶变换在纹理图像的分析中具有许多优点，如具有明确的物理意义、其功率谱具有平移不变性、有快速算法等等。因此，在纹理分析中，常借助于傅里叶频谱的频率特性来描述周期的或近乎周期的二维图像模式的方向性。这里常借用傅里叶频谱的三个性质：

- ① 傅里叶频谱中突起的峰值对应纹理模式的主方向；
- ② 这些峰在频域平面的位置对应模式的基本周期；
- ③ 如果利用滤波把周期性成分除去，剩下的非周期性部分将可用统计法描述。

一个大小为  $M \times N$  的图像  $I$  的离散傅里叶变换为

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-j2\pi(\frac{mu}{M} + \frac{nv}{N})} \quad (5-80)$$

式中， $u = 0, 1, 2, \dots, M-1, v = 0, 1, 2, \dots, N-1$ ， $f(m, n)$  表示图像位于  $(m, n)$  处的像素灰度值，同样，给定  $F(u, v)$ ，可以通过傅里叶逆变换获得  $f(m, n)$ ，即

$$f(m, n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi(\frac{um}{M} + \frac{vn}{N})} \quad (5-81)$$

式中， $m = 0, 1, 2, \dots, M-1, n = 0, 1, 2, \dots, N-1$ 。

利用傅里叶变换进行图像的纹理分析，一般先对图像（或图像中的一个窗口）进行傅里叶变换后计算傅里叶功率谱，并利用功率谱表示图像的纹理特征。由于功率谱是按频率统计图像的纹理能量的，因此功率谱反映了纹理的频率分布。一个信号的傅里叶频谱包括两部分，一是幅度特性，一是相位特性。

令  $F(u, v) = R(u, v) + jI(u, v)$ ，则有

$$\text{幅度: } |F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)} \quad (5-82)$$

$$\text{相位: } |\phi(u, v)| = \tan^{-1} \left( \frac{I(u, v)}{R(u, v)} \right) \quad (5-83)$$

$$\text{能量谱: } |E(u, v)| = R^2(u, v) + I^2(u, v) = |F(u, v)|^2 \quad (5-84)$$

频谱是图像的重要特征，反映图像的灰度分布。如果图像中目标形状或排列呈现方向性，那么具有较高值的频谱也会呈现出与图像目标方向正交的方向性分布。具体在纹理图像分类中，纹理特征一般采用傅里叶能量谱的前几个系数（如 10 个或 20 个，依具体情况而定）来代表。因为图像的主要内容主要集中在能量谱的低、中频段，高频能量较少，并且人眼对其也不太敏感，因此完全可以用傅里叶能量谱的前几个系数来表示图像的纹理特征，并对纹理图像进行分类。

除了直接利用傅里叶能量谱系数来表示图像的纹理特征外，还可以在能量谱的基础上进行特征的二次提取，以此来进一步判别图像纹理的粗细、方向和综合特征等。如果纹理具有空间周期性，或具有确定的方向性，则能量谱在对应的频率处会有峰。以这些峰为基础可组建纹理特征，常采用的楔环特征原理如图 5.21 所示。

#### (1) 环状采样

环状采样体现了纹理的粗糙度及周期性。设

$$P(u, v) = |F(u, v)|^2, \quad u, v = 0, 1, \dots, N-1 \quad (5-85)$$

将  $(u, v)$  直角坐标系转换成  $(r, \varphi)$  极坐标系, 得到  $P(r_i, \varphi_i)$ 。

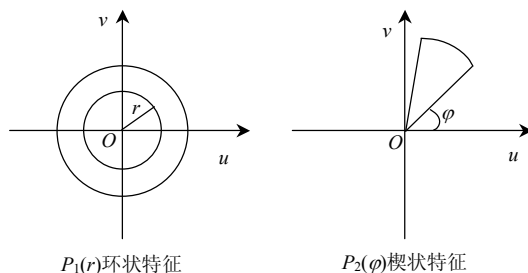


图 5.21 频谱的楔环特征

固定  $r$  对  $\varphi$  求和, 得到

$$P_1(r) = \sum_{i=1}^n P(r, \varphi_i), \quad r = r_1, r_2, \dots, r_n \quad (5-86)$$

称  $P_1(r)$  为环特征。对于不同的  $r, \varphi$  的求和路线形成环形。

$P_1(r)$  揭示了不同频率分量的能量分布信息, 在纹理较粗的情况下, 若  $r$  较小,  $P_1(r)$  很大,  $r$  很大时,  $P_1(r)$  反而较小; 而在纹理较细的情况下,  $r$  变化对  $P_1(r)$  的影响不是很大。

## (2) 楔状采样

楔状采样体现了纹理的方向信息。设

$$P_2(\varphi) = \sum_{i=1}^n P(r_i, \varphi), \quad \varphi = \varphi_1, \varphi_2, \dots, \varphi_n \quad (5-87)$$

称  $P_2(\varphi)$  为楔特性。对不同的  $\varphi, r$  的求和路线形成楔状。

分析由  $P_2(\varphi)$  构成的直方图, 当某一纹理图像沿  $\theta$  方向的线、边缘等大量存在时, 则在频率域内沿  $\theta + \frac{\pi}{2}$ , 即与  $\theta$  角方向成直角的方向上能量出现的比较集中。如果纹理不表现出方向性, 则频率谱也不呈现方向性。

根据傅里叶变换的幅度可以判断图像的相关性, 若两幅图像幅度相等, 则其完全相关这一规律, 参考文献(杨育彬 2003)提出了一种基于傅里叶变换的纹理分析法。其基本思想是, 对图像进行傅里叶变换, 得到图像的能量谱, 构造能量谱半径直方图和能量谱角度直方图。

对于式 (5-83), 令  $U = 2\pi \cdot \omega_1, V = 2\pi \cdot \omega_2$ , 再对  $\omega_1, \omega_2$  进行极坐标变换, 即

$$\begin{cases} \rho = \sqrt{\omega_1^2 + \omega_2^2} \\ \theta = \arctg(\frac{\omega_2}{\omega_1}) \end{cases} \quad (5-88)$$

式中,  $\rho$  为极坐标半径,  $\theta$  为相角大小。

首先在不考虑角度  $\theta$  的变化, 对二维频谱在环形区域内产生一个与方向无关的一维频域半径函数, 定义如下:

$$f(\rho) = \sum_{\theta=0}^{2\pi} |F(\rho, \theta)|^2 \quad (5-89)$$

$f(\rho)$  函数中峰值点的大小反映了图像中纹理基元的信息, 峰值点越小, 图像的纹理基元越大; 峰值点越大, 其纹理基元越小; 如果峰值点的数目很多或不明显, 则表示图像纹理杂乱无章。对于  $f(\rho)$  函数, 针对其对应的值点, 选择合适的量化级数, 对其直方图分布进行统计构造半径直方图。半径直方图表征了能量谱半径大小的分布频率

$$H = \left\{ \left\langle h[R_1], h[R_2], \dots, h[R_n] \right\rangle \mid \sum_{k=1}^n h[R_k] = 1, 0 \leq h[R_k] \leq 1 \right\} \quad (5-90)$$

式中,  $h[R_k] = \frac{\sum_{i=1}^m \begin{cases} 1, & \text{当 } R_{f_i(\rho)} = k \\ 0, & \text{其他} \end{cases}}{m}$  为  $R_k$  量级的值点频数,  $f_i(\rho)$  为  $f(\rho)$  函数中所有的取值  $1 \leq i \leq m$ ,  $R_{f_i(\rho)}$  为值点  $f_i(\rho)$  ( $1 \leq i \leq m$ ) 对应的量化级。

其次, 考虑能量谱角度  $\theta$  的分布情况, 将其作为纹理的方向性度量。不考虑  $\rho$  的变化, 对频谱在径向断面内取平均, 获得一个仅仅显示纹理特征方向特性的角度函数, 定义如下:

$$f(\theta) = \sum_{\rho=0}^1 |F(\rho, \theta)|^2 \quad (5-91)$$

$f(\theta)$  函数的峰值代表了纹理延伸的方向。如果峰值点过多或无明显峰值点, 则说明纹理无方向性或排列杂乱。对于  $f(\theta)$  函数, 采用频谱角度直方图对其分布进行统计。

#### 5.4.2 贝塞尔-傅里叶变换法

贝塞尔-傅里叶变换具有如下形式 (Beddow J.K. 1997)

$$G(R, \theta) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (A_{m,n} \cos m\theta + B_{m,n} \sin m\theta) J_m \left( Z_{m,n} \frac{R}{R_v} \right) \quad (5-92)$$

式中,  $G(R, \theta)$  是灰度函数 ( $\theta$  为角度);  $A_{m,n}, B_{m,n}$  是贝塞尔-傅里叶系数;  $J_m$  是第一种第  $m$  阶贝塞尔函数的零根 (zero root);  $R_v$  是视场半径。

采用这种方法可得到以下重要的纹理特征 (章毓晋 2007)。

- (1) 贝塞尔-傅里叶系数
- (2) 灰度分布函数 (灰度直方图) 的矩
- (3) 部分旋转对称系数

纹理由离散的灰度构成, 一个  $R$ -重 ( $R$ -fold) 对称操作可以通过将在  $G(R, \theta)$  的灰度与在  $G(R, \theta + \Delta\theta)$  的灰度比较来完成。由此可得到纹理的部分旋转对称系数 (Partial Rotational Symmetry Index) 如下:

$$C_R = \frac{\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (H_{m,n} R^2 \cos m \left( \frac{2\pi}{R} \right)) J_m^2(Z_{m,n})}{\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (H_{m,n} R^2) J_m^2(Z_{m,n})} \quad (5-93)$$

式中,  $R=1, 2, \dots$ ;  $H_{m,n} R^2 = A_{m,n} R^2 + B_{m,n} R^2$ 。

- (4) 部分平移对称系数

当将灰度沿半径对比 (如将  $G(R, \theta)$  与  $G(R, \theta + \Delta\theta)$  进行对比) 时就可发现部分平移对称性质。纹理的部分平移对称系数定义如下:

$$C_T = \frac{\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} H_{m,n}^2 J_m^2(Z_{m,n}) - [A_{m,n} A_{m-1,n} + B_{m,n} B_{m-1,n}] J_m^2(Z_{m-1,n}) \frac{\Delta R}{2R_y}}{2 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} H_{m,n}^2 J_m^2(Z_{m,n})} \quad (5-94)$$

它满足  $0 < C_T < 1$ 。

#### (5) 粗糙度

粗糙度 (coarseness) 定义为围绕一个像素  $(x, y)$  的 4 个邻域像素间的灰度差。分析表明粗糙度与部分旋转对称系数和部分平移对称系数有如下关系:

$$F_{crs} = 4 - 2(C_R + C_T) \quad (5-95)$$

#### (6) 对比度

当一些变量的值都分布在这些值的均值附近时, 称这种分布有较大的峰态 (kurtosis)。对比度 (contrast) 可借助峰态定义为

$$F_{con} = \mu^4 / \sigma^4 \quad (5-96)$$

式中,  $\mu^4$  是灰度分布模式关于均值的 4 阶矩,  $\sigma^2$  是方差。

#### (7) 不平整度

不平整度 (roughness) 与粗糙度和对比度有如下关系

$$F_{rou} = F_{crs} + F_{con} \quad (5-97)$$

#### (8) 规则性

规则性 (regularity) 是纹理元素在图像中变化的函数, 可定义为

$$F_{reg} = \sum_{t=1}^m C_R + \sum_{t=1}^n C_T \quad (5-98)$$

一幅具有高度旋转对称和高度平移对称的图像具有大的规则性。

### 5.4.3 小波变换法

近年来, 小波变换 (Wavelet transform) 成为人们研究的热点 (Chang T. et al 1993, Laine A. et al 1993)。由于小波变换对图像边界和奇异点的检测十分有效, 所以特别适用于图像纹理特征的提取。例如, 参考文献作者 (Remash N. et al 1995) 等人先对图像进行分割, 再提取边界上的奇异点, 对其邻近边界作参数曲线拟合, 然后对参数曲线进行小波变换, 利用变换后的系数构成特征向量进行图像检索, 取得了较满意的效果。Stack (Stack H.G. et al 1995) 等人则利用小波变换提取多尺度特征向量, 进行医学图像检索, 也获得不错的效果。

小波变换指的是将信号分解为一系列的基本函数  $\psi_{mn}(x)$ 。这些基本函数可以通过母函数  $\psi(x)$  的变形得到, 如下所示:

$$\psi_{mn}(x) = 2^{-m/2} \varphi(2^{-m}x - n) \quad (5-99)$$

式中,  $m$  和  $n$  是整数。这样, 信号  $f(x)$  可以表达为

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x) \quad (5-100)$$

为了将小波变换应用于图像处理, 需要有二维小波函数和尺度函数。采用可分离变量的方法可以由一维小波函数和尺度函数构造所需要的二维小波变换, 即

$$\begin{aligned}
 \phi(x, y) &= \phi(x)\phi(y) \\
 \psi^H(x, y) &= \psi(x)\phi(y) \\
 \psi^V(x, y) &= \phi(x)\psi(y) \\
 \psi^D(x, y) &= \psi(x)\psi(y)
 \end{aligned}
 \tag{5-101}$$

式中,  $\phi$  是二维尺度函数,  $\psi^H, \psi^V, \psi^D$  是3个二维小波函数。每个小波上的  $H$  表示水平方向,  $V$  表示垂直方向,  $D$  表示对角线方向, 由式 (5-101) 给出的尺度函数和小波函数, 可以定义一个伸缩和平移的基函数, 即

$$\begin{aligned}
 \phi_{j,m,n}(x, y) &= 2^{j/2} \phi(2^j x - m, 2^j y - n) = \phi_{j,m}(x) \phi_{j,n}(y) \\
 \psi_{j,m,n}^H(x, y) &= 2^{j/2} \psi^H(2^j x - m, 2^j y - n) = \psi_{j,m}(x) \phi_{j,n}(y) \\
 \psi_{j,m,n}^V(x, y) &= 2^{j/2} \psi^V(2^j x - m, 2^j y - n) = \phi_{j,m}(x) \psi_{j,n}(y) \\
 \psi_{j,m,n}^D(x, y) &= 2^{j/2} \psi^D(2^j x - m, 2^j y - n) = \psi_{j,m}(x) \psi_{j,n}(y)
 \end{aligned}
 \tag{5-102}$$

利用这些基函数可对图像沿行向 (水平方向) 和列向 (垂直方向) 进行滤波和2-下采样, 如图5.22所示。首先对原始图像  $I(x, y)$  沿行向 (水平方向) 进行滤波和2-下采样, 得到系数矩阵  $I_L(x, y)$  和  $I_H(x, y)$ , 然后再对  $I_L(x, y)$  和  $I_H(x, y)$  分别沿列向 (垂直方向) 滤波和2-下采样, 得到一层小波分解的4个子图  $I_{LL}, I_{HL}, I_{LH}, I_{HH}$ 。其中,  $I_{LL}$  为  $I(x, y)$  的 (粗) 逼近子图, 是原始图像在低分辨率上的近似,  $I_{HL}$  是  $I(x, y)$  的水平方向细节子图,  $I_{LH}$  是  $I(x, y)$  的垂直方向细节子图,  $I_{HH}$  是  $I(x, y)$  的对角线方向细节子图。对逼近子图  $I_{LL}$  重复此分解过程, 直到确定的尺度, 就构成了多尺度的小波分解。图5.23是图像二层小波分解的示意图。

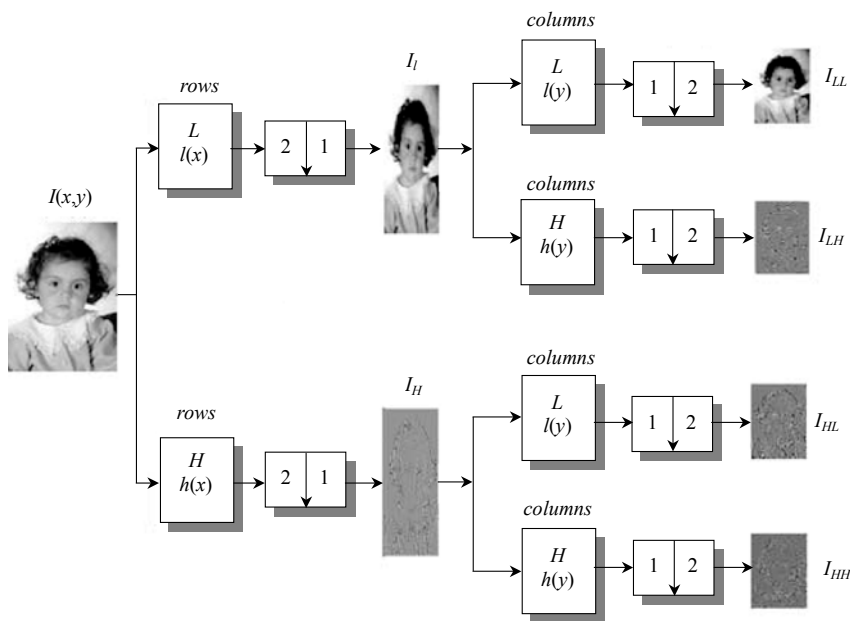


图 5.22 图像二层小波分解示意图



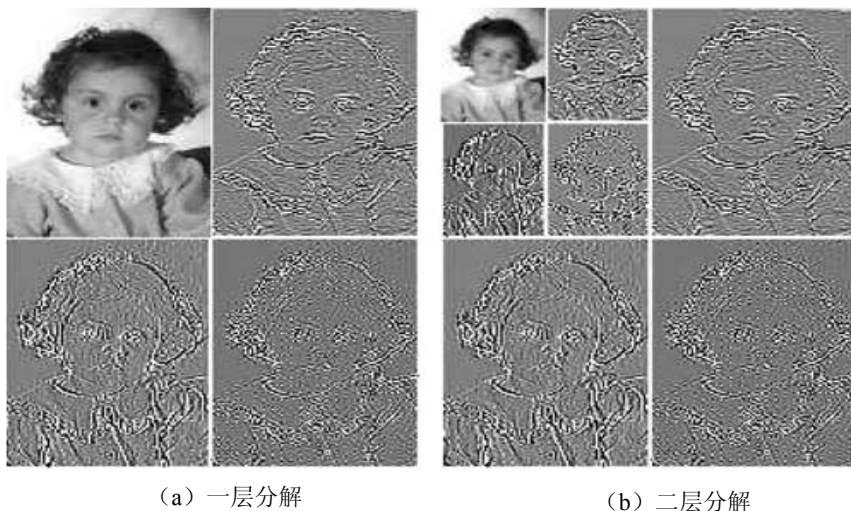


图 5.23 图像的多尺度分解

小波变换表示的纹理特征可以用每个分解层上能量分布的均值和标准方差来表示。假设局部纹理区域具有空间一致性,则变换系数的均值 $\mu_{mn}$ 和标准差 $\sigma_{mn}$ 可代表该区域用于分类和检索。均值 $\mu_{mn}$ 和标准差 $\sigma_{mn}$ 表示如下:

$$\begin{aligned}\mu_{mn} &= \iint |w_{mn}(x, y)| dx dy \\ \sigma_{mn} &= \sqrt{\iint (|w_{mn}(x, y) - \mu_{mn}|)^2 dx dy}\end{aligned}\quad (5-103)$$

用 $\mu_{mn}$ 和差 $\sigma_{mn}$ 作为分量,可以构成特征向量 $f\{\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \dots, \mu_{S-1, k-1}, \sigma_{S-1, k-1}\}$ 用于检索。

现有的小波变换,如Haar小波变换, Daubechies小波变换等,均存在两方面明显的局限性:其一,要求图像尺寸必须是2的幂;其二,图像超过边界时允许进行周期扩展。然而,为满足尺寸要求而进行的图像填补,会影响图像的特征,而周期扩展对局部图像显然不适合。针对以上情况, Dwong (Dwong M.K. et al 1995)等人提出了一种进行多分辨率分析的 $W$ 变换:先对图像中感兴趣的区域进行多尺度 $W$ 变换,然后统计经多尺度分解的特征图像的直方图,以此作为特征进行图像检索。不仅克服了以上两方面局限性,而且还具有光滑特性(尺度函数连续)。它既适合于各种尺寸的图像,也适合于对局部图像的变换分析;不仅可以用于整幅图像检索,也适用于大图像的目标检索。实验结果表明,该方法既实用又有效。

小波变换也常与其他技术相结合以获得更好的效果。Gross等人用小波变换、KL展开式和Kohomen多处理机系统来进行纹理分析(Gross M.H. et al 1995)。Thyagarajan等人则结合了统计法和变换法两者的优点,采用小波变换和共生矩阵来进行纹理分析(Thyagarajan K. et al 1994)。

针对小波分析技术在纹理特征提取方面的应用,近年来出现了很多研究成果。参考文献(吴均等 2001)基于不同纹理其信息从一个尺度到另一个尺度间的变化是不同的这样一个事实,构造了一个反映尺度依存关系的矩阵-尺度共生矩阵,在此基础上进行纹理特征的提取。参考文献(孟祥增 2004)采用Hough变换检测图像中的直线边沿,然后根据直线边沿的斜率分布来识别纹理图像中的条纹和网纹进而进行纹理图像的分类。参考文献(安志勇等 2007)提出一种基于Radon和小波变换的图像纹理特征检索算法,对原始图像进行坐

标系的旋转校正, 确定方向归一化图像, 再对方向归一化后的图像进行Radon变换, 根据Radon变换投影数据的几何特性, 构造适合投影数据的具有尺度和平移不变性的小波分解, 最后采用图像中各尺度小波系数的能量值作为图像的纹理特征。参考文献(安志勇等 2008)根据角向矩极大原理将检索图像进行坐标校正, 得到图像旋转不变的表示; 然后, 利用平移和尺度不变小波对检索图像进行分解, 得到具有平移、旋转和尺度不变的小波分解系数; 最后, 采用各尺度的小波能量值刻画图像的纹理特征。

#### 5.4.4 Gabor变换法

傅里叶变换是图像时域与频域分析的基础, 但以傅里叶变换为基础的经典分析方法只能得到图像时域或频域的整体信息, 无法刻画局部信息, 这对于非平稳信号则不能进行很好地分析。而 Gabor 变换能够很好地同时在时域和频域中兼顾对信号分析的分辨率的要求, 成为最常用的图像纹理特征提取方法。

Gabor 函数由 Gabor 博士于 1946 年提出, 通过高斯函数加上频移后产生。Gabor 滤波器是用 Gabor 函数作单位冲激响应的带通滤波器, 具有良好的滤波性能, 其输出可以看做输入信号的 Gabor 小波变换。生物学研究发现, 二维 Gabor 滤波器可以模拟生物的视觉系统, 能够很好地描述脊椎动物大脑初级视觉皮层部分的单细胞可接受信息域的分布, 因而在图像分析中具有重要的作用。而且, Gabor 滤波器在消除空域和频域二维联合不确定性方面是最优的, 它可以看成是方向、尺度可调的边界和直线检测器, 是被公认的信号表示尤其是图像辨识的最好方法之一, 所以可以通过 Gabor 滤波器检测出图像中不同方向和角度上的边缘和线条, 以提取图像中的纹理特征。

由于Gabor滤波可以看做是一种小波变换, 因此, 可以从小波变换的角度阐述该滤波器组的定义(王文惠 2001)。

设图像  $f(x, y)$ , 它的二维小波变换为

$$I_{mlpq} = \iint f(x, y) \varphi_{ml}(x - p\Delta x, y - q\Delta y) dx dy \quad (5-104)$$

$\Delta x$  和  $\Delta y$  是空间采样间隔, 通常设  $\Delta x = \Delta y = 1$ 。 $p$  和  $q$  是图像像素点的位置。 $m$  和  $l$  分别定义了小波变换的方向和尺度。即  $m = 0, \dots, M-1, l = 0, \dots, L-1$ 。其中,  $\varphi_{ml}(x, y)$  由小波变换的母波得到

$$\varphi_{ml}(x, y) = a^{-m} \tilde{\varphi}(\tilde{x}, \tilde{y}) \quad (5-105)$$

式(5-105)中,  $\tilde{x} = a^{-m}(x \cos \theta + y \sin \theta)$ ,  $\tilde{y} = a^{-m}(-x \sin \theta + y \cos \theta)$ 。

母波  $\varphi(x, y)$  通过  $a^{-m}$  尺度发生变化, 方向  $\theta$  变化的定义为

$$\begin{aligned} \theta &= l \Delta \theta \\ \Delta \theta &= 2\pi / L \end{aligned} \quad (5-106)$$

这个定义使得所有的滤波器具有同样的能量。当把Gabor函数作为母小波时, Gabor滤波就可以看成小波变换, 即

$$g(x, y) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W_x \right] \quad (5-107)$$

式(5-107)中  $W$  定义了滤波器的频率带宽。由神经物理学的研究,  $W = 0.5$ , 即取数字最高频率1的一半时, 完全符合人类的视觉系统(王文惠 2001)。

由 Gabor 小波的定义可以看出 Gabor 小波实质上是由一个复指数信号调制一个二维高斯函数得到的。设频域空间的坐标轴是  $u$  轴和  $v$  轴, Gabor 小波的频域特征是高斯函数的频谱沿着  $u$  轴平移得到

$$\begin{aligned}\psi(u, v) &= e^{-2\pi^2(\sigma_u^2 u^2 + \sigma_v^2 v^2)} * \delta(u - W) \\ &= e^{-2\pi^2(\sigma_u^2 (u-W)^2 + \sigma_v^2 v^2)} \\ &= e^{-\frac{1}{2} \left( \frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right)}\end{aligned}\quad (5-108)$$

式 (5-108) 中频率方差  $\sigma_u = \frac{1}{2\pi\sigma_x}$ ,  $\sigma_v = \frac{1}{2\pi\sigma_y}$ 。

在纹理分析中, 滤波器组的各个滤波器具有相同的能量, 但是在计算自然图像的相似性时, 由于自然图像的频谱幅度由低频到高频迅速下降, 导致 Gabor 滤波器的高频输出非常小。因此, 在计算小波变换的各个基函数时, 重新定义式 (5-105), 即

$$\hat{\varphi}_{ml}(x, y) = \tilde{\varphi}(\tilde{x}, \tilde{y}) \quad (5-109)$$

这样, 式 (5-104) 可以重新定义为

$$\hat{I}_{mlpq} = \iint f(x, y) \hat{\varphi}(x - p\Delta x, y - q\Delta y) dx dy \quad (5-110)$$

设图像的每一个像素点  $(p, q)$  经过滤波后, 得到的输出特征, 即

$$F_{mlpq} = \left| \hat{I}_{mlpq} \right| \quad m = 0, \dots, M-1 \quad l = 0, \dots, L-1 \quad (5-111)$$

输出特征值保留了能量信息, 忽略了位置信息。在提取图像的纹理特征时, 能量信息可以较好地反映图像的特征。

这样给定一幅图像后, 它的纹理特征可以由 Gabor 滤波器的输出得到, 即

$$E_{ml} = \frac{\sum_{p,q} F_{mlpq}^2}{\sum_{m,l,p,q} F_{mlpq}^2} \quad (5-112)$$

式 (5-112) 中的分母是标准化因子, 它的作用是增强图像内容, 减少图像在拍摄过程中光照度和对比度对图像特征的影响。

许多研究者针对 Gabor 变换也先后进行了深入研究。Jain A.K. (Jain A.K. et al 1991) 采用一组偶对称的 Gabor 滤波器, 对一组经过滤波后的图像逐步作线性变换  $\varphi(t) = \tanh(\alpha t)$ , 使计算的特征更具有特殊性。Gabor 特征的选择是基于局部能量意义上的, 而滤波器的选择是根据不同方向和频率通道上滤波图像功率和决定的, Jain 经过实验研究, 从 20 个滤波器中选出了 11~13 个滤波器对 Brodatz 纹理图像进行了分割, 取得了较好的分割结果。Alexandrov A.D. (Alexandrov A.D. et al 1995) 等人, 利用 120 个 Gabor 滤波器 (10 种尺度和 12 种方向) 输出能量的均值和标准差, 构成 240 维的纹理特征向量用于图像检索, 获得较理想的效果, 但该方法提取的特征向量维数偏高, 计算量太大。针对这一缺陷, Manjunath B.S. (Manjunath B.S. et al 1996) 等人提出了改进方法, 考虑到 Gabor 滤波函数是完备的非正交函数集, 他们首先消除系列 Gabor 滤波器的冗余度 (相关性), 然后又设计了一种自适应滤波器选择方法, 大大降低了计算量, 最后以 Gabor 小波变换系数的均值和方差作为纹理特征进行图像检索。

## 5.5 模型法纹理分析

基于模型的纹理分析是以图像的构造模型为基础,通过模型参数来定义纹理,模型的参数决定纹理的质量。模型法的主要问题是估计模型参数,使其所表示的纹理图像逼近原纹理图像。典型的方法是随机场模型法,如马尔可夫(Markov)随机场(MRF)模型法和 Gibbs 随机场模型法。

### 5.5.1 马尔可夫随机场模型法

马尔可夫随机场(MRF)是图像建模的重要工具。采用 MRF 来描述纹理图像时,首先需要对图像进行分块,在每个分块中采用最大似然估计(为计算方便经常采用伪似然估计)和最小二乘估计等方法估计模型参数,然后对一系列的模型参数进行聚类,确定该像素点及其邻域情况下该像素点最可能归属的概率。近 20 年来,由于 MRF 模型在理论上可以产生任何模式的纹理,所以,利用马尔可夫随机场模型法来描述纹理特征取得了很大的成功。这一模型假设每个像元的密度与邻域像元有关,与其他像元无关,紧靠的元素有直接交互作用,另外,全局的影响也可以传播。在纹理分析中,Markov 随机场模型基于纹理满足随机、静态等前提条件。

对于一个给定的图像  $\{y_0(i, j)\}$ , 令  $y = y_0 - \mu$ ,  $\mu$  为  $y_0$  的均值,则零均值输出  $\{y(i, j)\}$  满足下列差分方程(孙即祥等 2001)

$$y(i, j) = \sum_{(k, l) \in \eta} g_{k, l} y(i - k, j - l) + e(i, j) \quad (5-113)$$

式中,  $\{g_{k, l} | (k, l) \in \eta\}$  为 MRF 模型参数,在这里为待估量,  $\eta$  为  $(i, j)$  的邻域(指标集);  $\{e(i, j)\}$  为零均值的高斯噪声场,其相关结构为

$$E[e(i, j)e(m, n)] = \begin{cases} \sigma_e^2, & (i, j) = (m, n) \\ -g_{i-m, j-n} \sigma_e^2, & (m, n) \in \eta \\ 0, & \text{其他} \end{cases} \quad (5-114)$$

$$E[y(i, j)e(m, n)] = \begin{cases} \sigma_e^2, & (i, j) = (m, n) \\ 0, & \text{其他} \end{cases}$$

对 MRF, 有  $g_{k, l} = g_{-k, -l}$ , 从而 MRF 纹理模型可表示为

$$y(i, j) = \sum_{(k, l) \in \eta_s} g_{k, l} [y(i - k, j - l) + y(i + k, j + l)] + e(i, j) \quad (5-115)$$

这里  $\eta_s$  为非对称域, 它满足:

$$\left. \begin{aligned} \eta &= \eta_s \cup \overline{\eta_s} \\ \eta_s \cap \overline{\eta_s} &= \Phi \\ \overline{\eta_s} &= \{(-k, -l) | (k, l) \in \eta_s\} \end{aligned} \right\} \quad (5-116)$$

记点  $(i, j)$  为  $s$ ,  $\{y(i - k, j - l) + y(i + k, j + l)\}$  和  $\{g_{k, l}\}$  分别对应写出矢量形式  $Z_s$  和  $g$ , 则式(5-115)可以写为

$$y_s = g^T z_s + e_s \quad (5-117)$$

则 MRF 模型参数的最小二乘估计为

$$\begin{aligned}\hat{g} &= \left( \sum_{s \in G} z_s z_s^T \right)^{-1} \left( \sum_{s \in G} y_s z_s \right) \\ \hat{\sigma}_e^2 &= \frac{1}{n_G} \sum_{s \in G} \left( y_s - z_s^T \hat{g} \right)^2\end{aligned}\quad (5-118)$$

式中,  $G$  为数据窗;  $n_G$  为该区域中像素的个数。

马尔可夫随机场模型能很好地逼近纹理, 比较充分地反映出纹理图像局部上下文信息, 以它为基础衍生出了众多纹理表达合成的算法。Rama Chellappa 提出了一种马尔可夫随机场模型参数估计方法 (Rama Chellappa. 1985), 并根据这些参数通过高斯马尔可夫随机场采样合成纹理。该算法关注的是邻域相关的纹理局部特征, 对于那些全局特征与全局特征相一致的随机纹理刻画较好, 而对于结构纹理和混合纹理的全局特征很难描述。以 Efros 和 Leung 的无参采样纹理合成算法为标志 (Efros A. A. et al 1999), 产生了许多以 MRF 模型为基础的算法, 通过在样图中搜索一定邻域内的匹配点逐步合成出结果。

在实际应用中, MRF 模型对于描述在微观结构上一致的图像纹理具有较好的效果, 诸如草地、沙地或动物的羽毛等, 但对点状、条状或大尺度结构的纹理, 则需要额外的辅助方法加以描述。例如, 为弥补 MRF 模型的不足, 可将小波变换与之结合起来, 建立图像特征的多尺度随机场模型 (MSRF)。多尺度下的 Markov 随机场模型要比固定尺度下的马尔可夫随机场模型具有许多优越性。一方面, 多尺度下的马尔可夫随机场中的马尔可夫链结构是分级构成的, 在模型中无需人为地赋予每个像素点在空间域中的阶数, 较容易实现参数的估计, 解决了 MRF 模型中的归一化常数难题。另一方面, 也可以利用参数同时控制粗、精尺度下的图像行为。因此, 多尺度下的马尔可夫随机场模型能够更好地精确描述图像行为。

### 5.5.2 Gibbs随机场模型法

一幅数字图像的局部特性反映了一个像素的灰度值仅与其邻域的灰度有关, 而这种局部特性可由马尔可夫随机场模型来加以刻画, 但由于 MRF 是用条件概率分布来定义的, 因此无法很好地反映随机场变量的联合概率分布, 使得其在解决全局特性方面存在各种困难。主要原因有以下几个方面:

- ① 条件概率密度函数必须满足某些相容条件, 而这些条件在图像中不易被校验;
- ② 联合概率分布不能直接从局部计算得到;
- ③ 一个状态局部空间特性和其参数间的关系不是显而易见的。

为此, Geman S 和 Geman D (Geman S. et al 1984) 在 1984 年强调了 Gibbs 分布与随机场过程的马尔可夫性质之间的等价性, 提出采用 Gibbs 分布来定义 MRF, 称为吉布斯随机场。Gibbs 随机场凭借其所提供的联合分布特点, 已经成为纹理特征描述中一种流行的模型。定义如下: 称随机场  $F$  为在位置集  $S$  上关于邻域系统  $N$  的吉布斯随机场, 当且仅当随机场的模型满足如下吉布斯分布形式:

$$P(f) = Z^{-1} \exp \left( -\frac{1}{T} U(f) \right) \quad (5-119)$$

式中,  $Z = \sum_{f \in F} \exp\left(-\frac{1}{T}U(f)\right)$  为归一化常数, 也称为分拆函数;  $U(f)$  称为能量函数, 其表达式为

$$U(f) = \sum_{c \in C} V_c(f) \quad (5-120)$$

### 5.5.3 自回归模型

自回归模型 (Simultaneous Auto-Regressive, SAR) 是马尔可夫随机场 (MRF) 模型的一种应用实例。在 SAR 模型中, 每个像素的强度被看做是随机变量, 可以通过与其相邻的像素一起来描述。

(1) 基本的 SAR 模型 (Khotanzad A. et al 1987)

基本的 SAR 模型, 描述了像元与其邻域像元的关系, 是有条件马尔可夫模型 (Conditional Markov Model, CMM) 的子集。假设有尺寸为  $M \times M$  的图像,  $s = (s_x, s_y)$  是其中任意像元,  $s_x, s_y \in [0, M-1]$ , 则像元  $s$  的 SAR 模型值定义为

$$F_{SAR}(s) = \mu + \sum_{r \in R} \theta(r) f(s+r) + \varepsilon(s) \quad (5-121)$$

式中,  $\mu = \frac{1}{M^2} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x,y)$  为图像亮度的均值,  $R$  为  $s$  点的邻域集, 通常表示为如式 (5-122)

所示的二阶邻域;  $\theta(r)$  为模型参数, 它描述了邻域内像元之间的依存关系, 一般取对称模型, 即  $\theta(r) = \theta(-r)$ ;  $\varepsilon(s)$  为具有方差  $\sigma^2$  的零均值高斯随机变量, 其方差决定了噪声级别。

$$\begin{array}{ccc} (-1,1) & (0,1) & (1,1) \\ (-1,0) & (0,0) & (1,0) \\ (-1,-1) & (0,-1) & (1,-1) \end{array} \quad (5-122)$$

从视觉方面考虑, SAR 模型同样可以反映纹理的粒度和粗糙度:  $\sigma$  可以度量纹理的粒度 (granularity or business), 高的方差表示好的粒度或低的粗糙度;  $\theta$  的变化可以提取不同的方向特性, 例如, 垂直方向纹理在  $\theta(0,1), \theta(0,-1)$  的条件下会有高的模型值。显然, SAR 模型的关键在于参数  $\sigma$  和  $\theta$  的选取, 通常它们由最小平方误差法 (Least Square Error) 和极大似然估计 (Maximum Likelihood Estimation) 确定。

(2) 旋转不变 SAR 模型

由于基本 SAR 模型不满足旋转不变性, 参考文献 (Mao J. et al 1992) 等人基于基本 SAR 模型提出了旋转不变 SAR 模型 (RISAR)。与基本 SAR 模型不同的是, RISAR 中像元的邻域定义采用了圆形, 即邻域定义为当前像元为圆心的半径不同的  $K$  个同心圆。像元的模型值基于邻域像元的加权和计算, 即

$$F_{RISAR}(s) = \mu + \sum_{i=1}^K \theta_i f_i(s) + \varepsilon(s) \quad (5-123)$$

$$f_i(s) = \frac{1}{8i} \sum_{r \in N_i} \omega_i(r) f(s+r) \quad (5-124)$$

式中,  $K$  为描述圆邻域的同心圆个数,  $N_i$  为像元  $s$  第  $i$  个同心圆的像元个数;  $\omega_i$  为事先定义好的第  $i$  个同心圆上像元的权值。其余参数的含义同基本 SAR 模型。

### 5.5.4 多尺度自回归模型

定义合适的 SAR 模型需要确定相邻像素集合的范围。然而,固定大小的相邻像素集合范围无法很好地表达各种纹理特征。为此, Mao (Mao J. et al 1992) 等人提出多分辨率的自回归纹理模型 (Multi-Resolution SAR, MRSAR), MRSAR 能够在多个不同的相邻像素集合范围下计算纹理特征。参考文献 (Picard R.W. et al 1993, Manjunath B.S. et al 1996) 中给出了 MRSAR 纹理特征和其他纹理特征对于图像检索的性能对比。实验结果表明 MRSAR 能够较好地识别出图像中各种纹理特征。

多尺度 (多分辨率) 方法是一种双随机模型 (Luo J. et al 2001), 它基于如下两种假设:

① 假设  $Y$  中每个观测到的像素点的统计行为完全信赖于  $X$  中一个相应的非观察标称的像素点, 每个标称值指定  $M$  中具有不同统计行为类型中的一种。

② 假设由粗尺度到精尺度下的随机场序列是一个马尔可夫链, 即  $X^{(n)}$  的分布仅信赖于  $X^{(n+1)}$  的分布。采用连续 MAP 估计方法, 得到  $X^{(n)}$  的递推公式如下:

$$\hat{x}^{(L)} = \arg \max_{x^{(n)}} \log P_{y/x^{(L)}}(y/x^{(L)}) \quad (5-125)$$

$$\hat{x}^{(n)} = \arg \max_{x^{(n)}} \left\{ \log P_{y/x^{(n)}}(y/x^{(n)}) + \log P_{x^{(n)}/x^{(n+1)}}\left(x^{(n)}/\hat{x}^{(n+1)}\right) \right\} \quad (5-126)$$

式 (5-126) 中, 第一个式子为初始条件,  $L$  为最大尺度值。在上述递推公式中, 第一部分给出了对尺度  $n$  下的标称值, 表示观测数据  $y$  的似然率, 第二部分表示  $X$  随机变量行为的先验信息。因此, 这部分使得这种分割方法适合于处理具有较大纹理和平滑边界的图像, 弥补了传统 MRF 模型的不足。

### 5.5.5 基于分形模型

传统的几何学认为任何形状的维数都是整数, 如点是一维, 线是一维, 平面是二维, 立体是三维, 而分形几何学提出了分形维的概念, 认为很多几何形状均可能具有分形维数, 而这个分形维数成为应用分形理论研究的主要参数。1975 年, Mandelbrot 提出了用分形来表达那些没有特征长度, 具有无限精细结构的图形、构造及现象, 从而来描述事物的自相似性。Mandelbort 曾经分别下过如下两个定义。

①  $D(A) > d(A)$  的集合  $A$ , 称为分形集, 其中  $D(A)$  为集合  $A$  的分形维数,  $d(A)$  为其拓扑维数。一般来说,  $D(A)$  不是整数, 而是分数。

② 部分与整体以某种形式相似的形, 称为分形。然而经过理论与应用的检验, 人们发现这两种定义很难包括分形如此丰富的内容。而实际上, 到目前为止, 分形仍然没有一个确切的定义。流行的对分形的描述如下。

一般地, 如果  $F$  上具有以下性质的集合, 则称它是一个分形:

- ①  $F$  具有精细的结构, 即有任意小比例的细节;
- ②  $F$  是如此不规则, 以至于它的整体和局部都不能用传统的几何语言来描述;
- ③  $F$  通常具有某种自相似形式, 可能是近似的或统计的;
- ④ 一般地,  $F$  的“分形维数”大于它的拓扑维数;
- ⑤ 在大多数令人感兴趣的情形下,  $F$  可以以某种非常简单的方法产生, 可能由迭代

产生。

对于各种不同的分形,有的同时具有上述的全部性质,有的可能只有大部分性质,而对某个性质有例外,但这并不影响我们把这个集合称为分形。

对于图像中分形的研究,较早的有 Pentland (Pentland P.1984),他通过对自然景物图像纹理的研究,证明大多数图像纹理表面的粗糙度与分形模型很接近,自然界的物体(如云雾、青山、凹凸不平的地面、风化而斑驳的岩石)大多具有比较强的分形特征,分形模型在一定的尺度范围内可以很好地与自然背景的表面和实际结构相吻合。这就使我们可以利用各种不同的特性参数(分形特性)来区分不同的物体。分形的特殊性使其不能用普通的面积、体积、导数等概念来对其进行描述,而需要寻找一类新的研究和描述工具。分形维数是其中最重要也是最常用的工具,特别是在图像处理领域,主要原因有以下几个方面:

- ① 分形维数是一种可以用实验手段定量估计的数值参数;
- ② 图像分形维数的大小与人眼对物体表面的粗糙度大小的感受密切相关;
- ③ 分形维数的大小对图像的尺度和线性变换具有稳定性。

分形维数是分形对象极其重要的特征,它可以对分形对象的复杂度、不规则程度和全局正则性进行定量描述。分形维数在图像处理中的应用是以两点为基础的:一是自然界中不同种类的形态物质一般具有不同的分形维;二是自然界中的分形与图像的灰度表示之间有着一定的对应关系。准确地估计分形对象的分形维数在分形信号的处理和分析中非常关键。目前常用的分形维数有相似维数、Hausdorff 维数、计盒维数等,其中计盒维数由于计算相对简单得到了广泛的应用。

设  $F$  是  $R^n$  中任一非空有界子集,  $N_\delta(F)$  是直径最大为  $\delta$ , 可以覆盖  $F$  的集的最少个数, 则  $F$  的上、下盒维数分别定义为 (李永根等 2002):

$$\begin{aligned}\overline{\text{DIM}}_B(F) &= \lim_{\delta \rightarrow 0} \frac{\log_2 N_\delta(F)}{-\log_2 \delta} \\ \underline{\text{DIM}}_B(F) &= \lim_{\delta \rightarrow 0} \frac{\log_2 N_\delta(F)}{-\log_2 \delta}\end{aligned}\quad (5-127)$$

如果这两个值相等, 则称这共同的值为盒维数, 记为

$$\text{DIM}_B(F) = \lim_{\delta \rightarrow 0} \frac{\log_2 N_\delta(F)}{-\log_2 \delta} \quad (5-128)$$

计算图像分形维数的方法很多, 常见的有自相似维、容量维数、Hausdorff 维数、谱维数、拓扑维数、布朗维数、盒子维数、信息维数等, 大多数和定义都是基于“尺度  $r$  下的度量”这一思想: 对每个  $r$ , 在不考虑尺度小于  $r$  的不规则集下度量一个集合  $A$ , 观察在  $r \rightarrow 0$  时这些度量的变化。集合的维数由  $r \rightarrow 0$  时度量  $N_r$  所服从的指数率确定。在实际应用中, 针对不同的研究对象采取不同的维数定义方法。比较实用的从速度和精度上考虑, 有以下几种:

#### (1) 计盒法 (Chaudhuri B.B. et al 1995)

将灰度图像中像素点的坐标考虑为三维空间的  $X, Y$  坐标, 而各像素点的灰度为另一坐标  $Z$ 。假设有一个单位  $m$  维体正好将研究对象覆盖, 则线度  $\delta$  可以视为相对于这个单位  $m$  维体的尺度压缩因子。为保证研究对象在  $X, Y, Z$  3 个方向上有相同的压缩因子, 对盒子的尺寸作如下限制



$$\frac{G}{g} = \frac{W}{w} = \frac{L}{l} = \frac{1}{r} \quad (5-129)$$

式中,  $l, w, g$  分别为盒子的长, 宽, 高;  $G$  为图像中灰度级数;  $L$  和  $W$  分别为图像的长度和宽度;  $r$  为比例因子。用这样的盒子来覆盖三维空间中的图像表面, 对应于不同的  $r$ , 得到所需要的盒子数为  $N(r)$ , 这样就得到不同的  $(r, N(r))$  数据对, 盒维数  $\text{DIM}_B F$  即为  $(r, N(r))$  在双对数坐标上的斜率。令  $x_r = -\log_2 r, y_r = \log_2 N(r)$ , 用最小二乘法估计分形维数  $\text{DIM}_B(F)$  为

$$\text{DIM}_B(F) = \frac{\prod_r (x_r - \overline{x_r})(y_r - \overline{y_r})}{\prod_r (x_r - \overline{x_r})(x_r - \overline{x_r})} \quad (5-130)$$

式中,  $\overline{x_r}, \overline{y_r}$  是  $x_r, y_r$  的统计平均值。

### (2) 功率谱法

对图像先作傅氏变换成为频谱图, 其功率谱为  $|P(w)|^2$ , 而频率半径为  $R = \sqrt{U^2 + V^2}$ , 做出功率谱与频率半径的双对数图, 根据线性回归法求取分形维数值。

### (3) 地毯覆盖法

设分形表面为  $g(i, j)$ , 形象地用厚度为  $2\delta$  的地毯覆盖, 则地毯的表面点集为  $t_\delta(i, j)$ , 下表面点集为  $b_\delta(i, j)$ , 初始状态为  $t_0(i, j) = b_0(i, j) = t(i, j)$ , 当厚度  $\delta = 1, 2, 3, \dots$  变化时

$$\begin{aligned} t_\delta(i, j) &= \max \left\{ t_{\delta-1}(i, j) + 1, \max_{(m,n) \in S} t_{\delta-1}(m, n) \right\} \\ b_\delta(i, j) &= \max \left\{ b_{\delta-1}(i, j) - 1, \max_{(m,n) \in S} b_{\delta-1}(m, n) \right\} \end{aligned} \quad (5-131)$$

式中,  $S$  为点  $(i, j)$  邻域点集, 则在尺度  $\delta$  下, 毯的面积为

$$A(\delta) = \left[ \sum_{i,j} (t_\delta(i, j) - b_\delta(i, j)) \right] / 2\delta \quad (5-132)$$

做出  $\delta$  与  $A(\delta)$  的双对数图, 根据线性回归法求取分形维数值。

### (4) 布朗分形法 (Pentlend P.1984)

设  $I(x)$  为一分形布朗函数, 对于任意的  $x$  和  $\Delta x$ , 则满足

$$P_r \left\{ \frac{I(x + \Delta x) - I(x)}{\|\Delta x\|^H} < t \right\} = F(t) \quad (5-133)$$

式中,  $x$  为  $N$  维欧式空间  $R^n$  中的任一点;  $F(t)$  为满足高斯分布  $N(0, \sigma^2)$  随机函数;  $P_r$  表示概率,  $\|\bullet\|$  表示范数 ( $\bullet$  表示  $\|\bullet\|$  中的表达式); 参数  $H$  为频谱指数,  $\Delta x$  为偏移矢量, 当  $H = 0.5$  且  $F(t)$  为  $N(0, 1)$  高斯分布时,  $I(x)$  为一般布朗运动。分形布朗运动是一般布朗运动的推广。其中,  $I(x)$  的分形维数  $D$  和  $H$  的关系满足  $D = N + 1 - H$ ,  $N$  为拓扑维数。对图像而言  $N = 2$ , 则有  $D = 3 - H$ 。

由  $I(x)$  的定义式可以证明其满足以下性质:

$$E \left[ I(x + \Delta x) - I(x) \right] = C \|\Delta x\|^H \quad (5-134)$$

式中,  $C$  为常数,  $E(\bullet)$  表示数学期望算子, 对式 (5-134) 两边取对数得

$$\log E \left[ I(x + \Delta x) - I(x) \right] - H \log \|\Delta x\| = \log C \quad (5-135)$$

因此,  $E[I(x + \Delta x) - I(x)]$  与  $\log \|\Delta x\|$  的对数图是一条直线, 斜率为  $H$ , 从而近似估计出分形维数。

近年来, 在实际的工程应用中, 研究者们针对一些分形维数的定义, 也提出了许多关于分形维数计算的方法, 如在图像处理方面有 Gangepain 等的计网格元法 (Reticular Cell Counting)、Kelle 等的基于概率的估算法、基于分形布朗运动自相似模型的估计法及 Chaudhuri B.B. 和 Sarkar N. 提出的微分计盒法 (Differential Box-Counting, DBC) 等。

在图像处理中, 由于相同维数的纹理看起来并不完全相同, 因此仅用分形维数来描述纹理区分能力还不够强。为此, 又引入凹陷  $A$  的概念, 它反映了具有相同分形维数但视觉效果不同的纹理的特性。

给定一个分形点集合  $D$ , 用  $P(n)$  代表有  $n$  个点在以  $D$  中任一点为中心的  $L \times L$  盒中的概率 ( $\sum_{n=1}^N P(n) = 1$ )。设  $N$  为盒中可能有的点数, 则凹陷为

$$A = \frac{S_2 - S^2}{S^2} \quad (5-136)$$

式中

$$\begin{aligned} S &= \sum_{n=1}^N nP(n) \\ S_2 &= \sum_{n=1}^N n^2 P(n) \end{aligned} \quad (5-137)$$

凹陷  $A$  代表了对纹理集合的一种二阶统计, 大的  $A$  代表了具有较大颗粒性的纹理, 而小的  $A$  代表了具有较小颗粒的纹理 (Wu C.M. et al 1992)。

分形维数反映的是图像的粗糙度, 不同的结构在其边界处灰度变化不再连续, 而是产生一个突变, 这在视觉上的感受正是粗糙度的变化, 也就是说利用分形的变化可以更好地反映边界的情况。基于对描述自然纹理的随机分形模型-分数 Brownian 运动 (FBM) 的指数  $H$  的估计, 还有一种基于纹理图像亮度差值的自相关函数来描述自然纹理的方法 (吴高洪等 2000)。

FBM 图集的分形维数与 FBM 的指数  $H$  存在着——对应的关系;  $D = 2 - H$ 。给定自然数  $s$  和  $t$ , 取  $s \neq t$ , 可将  $H$  表示为

$$H = \ln \left\{ \frac{R_t(s)}{R_t(t)} \right\} / \ln \left( \frac{s}{t} \right) \quad (5-138)$$

式中,  $R_t(s)$  和  $R_t(t)$  分别为  $s$  和  $t$  的自相关函数, 计算方法如下:

假设纹理图像  $T(x, y)$  的大小为  $M \times M$ ,  $0 \leq x, y \leq M$ , 首先计算原始纹理图像步长为  $s$  的亮度差值图像 (即 IFBM), 即

$$I(x, y) = B(x, y) - \frac{1}{\#N(s)} \sum_{k, l \in N(s)} B(x+k, y+l) \quad (5-139)$$

式中,  $N(s)$  是内外半径分别为  $s-1$  和  $s$  的圆环内像素的集合, 即

$$N(s) = \{k, l | (s-1)^2 < k^2 + l^2 \leq s^2\} \quad (5-140)$$

$\#N(s)$  为  $N(s)$  包含的像素数目。然后可由式 (5-141) 求得

$$R_t(s) = \frac{1}{M^2} \sum_{x,y=0^2}^{M-1} \left[ I(x,y) \frac{1}{\#N(s)} \sum_{k,l \in N(s)} I(x+k,y+l) \right] \quad (5-141)$$

在理想情况下,对于任意不同的一对数 $s$ 和 $t$ ,由式(5-141)计算出的FBM指数 $H$ 应该是相等的,而且有 $0 < H < 1$ 。但对于实际的纹理图像,对于不同的 $s$ 和 $t$ ,计算所得的 $H$ 不仅不相等,而且差别还很大。另外 $H$ 值也不局限于0和1之间。主要是由于:

- ① FBM仅是描述纹理图像的一种近似模型,实际的纹理图像并不完全符合FBM模型;
- ② 因为FBM模型并不是有限带宽的,数字抽样时有混叠现象的影响;
- ③ 数字图像成像过程非线性的影响。

从式(5-138)可以看出, $H$ 反映的是 $R_t(s)$ 和 $R_t(t)$ 之间的相对关系,而这种关系映射到原始的纹理图像上,就体现为纹理图像中定义在不同邻域上的局部统计特性之间的关系。可以认为 $H$ 随 $s$ 和 $t$ 变化的现象,正是相应纹理图像本身的局部统计特性的体现。因此,可将由不同的 $s$ 和 $t$ 计算而得的 $H$ 作为描述自然纹理的特征。对于细纹理图像,由于灰度分布比较随机,对于不同的 $s$ 和 $t$ ,其亮度差值图像的自相关函数值比较接近,由式(5-138)计算而得的各个 $H$ 也比较接近。相反,对于粗纹理,像素之间的相关性较强,对于不同的 $s$ 和 $t$ ,其亮度差值图像的自相关函数差别也比较大,因而各个 $H$ 的分布比较分散。由于可见, $H$ 可作为描述纹理粗糙度的特征。

### 5.5.6 Wold模型法

在图像分析中,马尔可夫随机场在图像建模方面取得了很大的成功,但也有很多缺点:一是对随机场的均匀性假设和实际中大多数图像不吻合,如果考虑非平稳随机场,会有大量的模型参数,计算量也随之增大;二是马尔可夫随机场缺乏先验知识往往不能满意地刻画图像的结构成分。基于Wold模型的随机场分割法则将图像随机场分解为确定性和不确定性两种类型,能更精确地刻画纹理的特点。

Wold方法最早被用于麻省理工学院媒体实验室开发的PHOTOBOOK系统中,它的基本原理是随机场分解Wold原理,即二维随机场可以被分解为三个正交的元素:谐波(harmonic)周期性分量、渐进为零(evanescent)的方向分量和一个非确定性(in-deterministic)的随机分量。Wold分解首先检测图像的谐波分量,并提取谐波峰值特征,再对剔除波分量后的随机场做二阶多分辨率自回归(MSAR)建模,提取模型参数和对应的协方差矩阵作为纹理特征。综合谐波特征和模型特征构成纹理的Wold描述。实验表明,Wold特征与人的感知特性符合较好,在针对Broadtz纹理图像的检索实验中取得了很好的效果(Picard R.W. et al 1993)。

Wold纹理特征描述法基于均匀二维随机场的分解(liu F. et al 1994)。描述如下:

设 $\{y(n,m)\}$ 是这样一个随机场,则它有唯一的正交分解,即

$$y(n,m) = w(n,m) + v(n,m) \quad (5-142)$$

式中, $\{w(n,m)\}$ 为非确定随机场(Indeterministic Random Field); $\{v(n,m)\}$ 为确定随机场(Deterministic Random Field)。 $\{v(n,m)\}$ 可以进一步分解为一族全序定义(Total-Order Definitions)的非对称半平面(Non-Symmetric Half-Plane),且其分割角度的斜率为有理数。

设 $O$ 为所有这些划分的集合,则 $\{v(n,m)\}$ 的分解可表示为

$$v(n, m) = p(n, m) + \sum_{(\alpha, \beta) \in O} e_{(\alpha, \beta)}(n, m) \quad (5-143)$$

式中,  $\{e_{(\alpha, \beta)}(n, m)\}$  是非对称半平面渐消失随机场,  $\{p(n, m)\}$  是半平面确定随机场。非确定随机场、半平面确定随机场(和谐随机场)、渐消失随机场就被称为随机场  $\{y(n, m)\}$  的 Wold 特征, 它们对应于纹理中的随机性、周期性、方向性等特征; 周期纹理具有较强的和谐随机场成分, 强方向性纹理的渐消失成分多, 而结构性弱的纹理的渐消失成分多, 而结构性弱的纹理的非确定随机场成分占主导地位, 这样, 随机场参数和纹理特性就联系在一起了。

## 5.6 纹理基元共生矩阵

灰度共生矩阵作为一种重要的基于纹理特征进行图像检索的方法, 具有特征提取和相似度计算简便的优点, 此方法在多数情况下是有效的, 但也存在着一些缺点(见 5.2.3 节)。为了解决利用灰度共生矩阵进行检索存在的问题, 结合纹理分析法中的结构分析法和统计分析法, 并借鉴方块编码的思想, 本节构造了一种纹理基元的共生矩阵(赵珊等 2006)。

### 5.6.1 方块编码算法

方块编码(Block Truncation Coding, BTC), 又称作方块截断编码, 是一种基于块的快速有损图像压缩技术。从本质上说, BTC 是一比特自适应矩保持量化器, 它在量化输出中预测输入图像块的某些统计特性, 其编码思想是: 把一幅图像划分成一系列不重叠的子图像块, 由于子图像块中各像素的灰度相近, 此时只选取两个适当的灰度, 子图像块中的任一像素灰度用这两个灰度中的某一个代替, 于是这个子图像块中的所有像素都映射成这两个灰度值表示的代码, 最后量化成二电平的输出, 这样在重建图像块中就可保持均值和第一个绝对中心矩特性。方块编码算法简单, 性能较高, 信道容错力较好, 计算负担较小, 对存储器要求较少, 重建图像质量较高, 这使其在实时图像传输方面得到很广泛的应用。其编码原理如下。

设选定的子图像块的两个代表灰度为  $a_0$  和  $a_1$ , 且  $a_1 > a_0$ , 子图像块中第  $i$  个像素的灰度为  $x_i$ , 对于该子图像块的灰度门限  $T$ , 像素  $x_i$  编码后的灰度可按式(5-144)计算, 即

$$Z_i = (1 - y_i)a_0 + y_ia_1 \quad (5-144)$$

式中,  $y_i = \begin{cases} 1; & x_i \geq T \\ 0; & x_i < T \end{cases}$

原子图像块的各像素的灰度  $\{x_1, x_2, \dots, x_m\}$ , 也可以说变换后的各像素的灰度  $\{z_1, z_2, \dots, z_m\}$  可以用  $\{a_0, a_1\}$  与  $\{y_1, y_2, \dots, y_m\}$  组合表示, 它们就是子图像的方块编码。由上述原理可知, 根据它们完全可以恢复出  $\{z_1, z_2, \dots, z_m\}$ 。

同时,  $a_0$  和  $a_1$  应具有与原子图像块中像素相近的灰度级,  $\{y_1, y_2, \dots, y_m\}$  只用  $m$  个比特。若  $a_0$  和  $a_1$  各用  $b_a$  比特, 则编码后每个像素的平均比特数为

$$\tilde{b} = (m + 2b_a) / m = 1 + 2b_a / m \quad (5-145)$$

一般灰度取 6~8 位, 如  $m = 4 \times 4$ , 由式(5-145)可知方块编码的压缩效果是显著的, 并且  $m$  越大,  $b$  越小, 但同时, 图像质量也随之下降, 通常  $m = 4 \times 4$ 。其中, 参数  $T$  选择通常遵守如下规则:

(1) 均方误差最小策略

取  $T = \frac{1}{m} \sum_{i=1}^m x_i$ , 子图块各像素灰度变换后的平方误差为

$$\varepsilon^2 = \sum_{i=1}^m (x_i - z_i)^2 = \sum_{x_i < T} (x_i - a_0)^2 + \sum_{x_i \geq T} (x_i - a_1)^2 \quad (5-146)$$

为求使  $\varepsilon^2$  取最小的  $a_0$  和  $a_1$ , 令  $\frac{\partial \varepsilon^2}{\partial a_i} = 0$ ,  $i = 0, 1$ , 可得

$$a_0 = \frac{1}{m-u} \sum_{x_i < T} x_i, \quad a_1 = \frac{1}{u} \sum_{x_i \geq T} x_i \quad (5-147)$$

式中,  $u = \sum_{i=1}^m y_i$ 。由式 (5-147) 可知,  $a_0$  和  $a_1$  分别是小于和不小于  $T$  的像素的灰度平均。

(2) 保持一, 二阶矩不变策略

仍然取  $T = \bar{x}$ ,  $u = \sum_{i=1}^m y_i$ , 保持子图像块一、二阶矩不变是指

$$\begin{aligned} m \bar{x} &= (m-u)a_0 + ua_1 \\ m \bar{x}^2 &= (m-u)a_0^2 + ua_1^2 \end{aligned} \quad (5-148)$$

由式 (5-148) 可解得

$$\begin{aligned} a_0 &= \bar{x} - \sigma [u/(m-u)]^{1/2} \\ a_1 &= \bar{x} + \sigma [(m-u)/u]^{1/2} \end{aligned} \quad (5-149)$$

式中,  $\sigma^2 = \bar{x}^2 - (\bar{x})^2$ 。

这种方案的一个改进是用  $\bar{x}$ 、 $\sigma^2$  代替  $a_0$  和  $a_1$  进行编码和传输, 这样可进一步减少计算量, 因为  $\sigma^2$  的动态范围更小些。

方块编码算法速度很快, 易于实现且计算量低, 它保持了重建图像的质量且保留了边缘信息。近年来不少图像检索领域的专家对其也进行了大量的研究 (Qiu Guoping et al 2003, Gerald Schaefer et al 2004, 赵珊等 2005, 赵珊等 2007a, 赵珊等 2007b)。

借鉴方块编码的思想, 结合人眼的视觉特性, 充分利用图像块中的边缘信息、形状信息及纹理信息来定义图像的纹理基元。针对纹理基元在图像中的分布特征, 结合灰度共生矩阵, 构造了一种基于纹理基元的共生矩阵方法, 用以描述图像的纹理特征。

## 5.6.2 纹理基元的提取

在提取纹理基元前, 针对彩色图像需要在特定的空间中进行一些预处理。

### 1. 图像的预处理

算法中, 我们利用像素间的色差来定义纹理基元, 要定量地表示这些颜色之间的差异只能在特定的颜色空间中进行, 算法采用了两种颜色处理的方法。

(1) 将彩色图像直接利用式 (5-150) 转化为灰度图像, 然后通过像素之间的灰度差来提取纹理基元。

$$I = 0.3R + 0.59G + 0.11B \quad (5-150)$$

这种方法不需要对彩色图像做过多的处理,简单方便,利用变换后的灰度值来表示原图像的颜色值。

(2) 在某种颜色空间中将颜色进行适当的量化,用量化后的颜色值代替像素的灰度值,通过计算像素间的颜色差来提取图像的纹理基元。

众所周知,对彩色图像的研究,必须在特定的颜色空间中进行。关于常用的颜色空间,我们在第3章已经详细介绍。一般认为,RGB颜色空间与人眼的感知差别很大,例如,距离为50的(0, 0, 0)与(50, 0, 0)两种RGB颜色被认为是同一黑色,而距离为50的(200, 150, 0)和(200, 200, 0)则是差别很大的两种颜色。由于HSV颜色空间直接对应于人眼色彩视觉特征的3要素,在色调(Hue)、饱和度(Saturation)和亮度(Value)3个分量中,色调尤其影响着人类的视觉判断,而且HSV空间各轴在视觉上彼此无关,从RGB到HSV的转换比较简单,在基于内容的图像检索中,应用这种颜色模型会更适合用户的视觉判断。因此,本节的算法采用HSV颜色空间作为对彩色图像进行处理的颜色空间。

对于图像颜色的量化,这里采用黄元元(黄元元 2003)提出的量化方法,将图像的颜色量化为36柄。量化过程如下:

$$H = \begin{cases} 0, & H \in [0^\circ, 60^\circ) \\ 1, & H \in [60^\circ, 120^\circ) \\ 2, & H \in [120^\circ, 180^\circ) \\ 3, & H \in [180^\circ, 240^\circ) \\ 4, & H \in [240^\circ, 300^\circ) \\ 5, & H \in [300^\circ, 360^\circ) \end{cases} \quad S = \begin{cases} 0, & S \in [0, 0.25) \\ 1, & S \in [0.25, 1] \end{cases} \quad V = \begin{cases} 0, & V \in [0, 0.3) \\ 1, & V \in [0.3, 0.8) \\ 2, & V \in [0.8, 1.0] \end{cases} \quad (5-151)$$

根据式(5-151)的划分,我们可以将很多虽然深浅不同但在视觉上仍属于同一类的颜色量化在同一区间内,使量化结果更加符合人类的视觉感受。

量化完成后,我们将HSV颜色空间划分为 $n_H \times n_S \times n_V$ 个区间,其中 $n_H$ 、 $n_S$ 和 $n_V$ 分别表示H、S和V三个分量的量化级数,即 $n_H = 6$ ,  $n_S = 2$ ,  $n_V = 3$ 。也就是说,我们将HSV空间划分为了36个区间。按照以上的量化级数,我们把H、S、V三个分量合并为一个一维特征向量,即

$$n = n_S \times n_V \times H + n_V \times S + I \quad (5-152)$$

因为这里 $n_S = 2$ ,  $n_V = 3$ , 从而式(5-152)又可写为

$$n = 6H + 2S + I \quad (5-153)$$

这样H、S、V三个分量在一维向量上分布开来,根据式(5-152), $n$ 的取值范围为[0, 35], 也就是说,可以用[0, 35]之间的任一个数来表示量化后图像中像素的颜色值。另外,为了后面算法描述的统一,对两种图像预处理后的像素的灰度值和颜色值统称为像素的颜色值,在此特别说明。

## 2. 纹理基元提取算法

图像的特征提取实际是寻求一种尽可能紧凑的特征码值来描述图像最本质的内容,而图像压缩编码的目的从本质来说就是对要处理的图像按一定的规则进行变换组合,尽可能地减少或消除图像中的冗余信息,从而以尽可能紧凑的代码来表征原图像。因此,从这个意义上来说,图像的某种压缩编码就可以看做是提取得到的图像特征。另外,每幅图像

的不同区域都有着不同的结构特征,有的区域灰度比较均匀,没有很明显的明暗对比,而有的区域却有着很复杂的灰度差,明暗对比明显。而且,人眼对灰度变化的敏感程度跟背景有关,它随平均灰度的变化而变化,即人眼对图像细节的分辨力与图像的灰度阶差有关。当图像本身的灰度阶差较小时,人眼的分辨力会降低,反之亦然。因此,可以将图像按照灰度阶差分成不同形状的块,来表示图像中的纹理信息。由此,以方块编码的思想为基础,根据图像块的灰度差来进行纹理基元的提取。

我们从 5.3 节纹理基元的定义可以看出,纹理基元包含了两个要素:其一,是该像素区域能表征图像的纹理分布;其二,是作为基元,研究的像素数目力求最小。综合这两个要素,算法中规定  $2 \times 2$  窗口为表征图像纹理的最小单位,称为纹理基元。具体提取算法如下:

假设  $I$  是一幅大小为  $M \times N$  的图像。首先将  $I$  划分为  $m \times m$  大小的互不重叠的子块,对于每个子块,按式 (5-154) 和式 (5-155) 计算块内像素的颜色均值  $\mu$  和平均颜色差  $\sigma$ ,即

$$\mu = \frac{\sum_{i,j} p(i,j)}{m^2} \quad (5-154)$$

$$\sigma = \frac{\sum_{i,j} \|p(i,j) - \mu\|}{m^2} \quad (5-155)$$

式中,  $p(i,j)$  表示位于  $(i,j)$  处的像素的颜色值。

按照方块编码的思想,在每个子块中,对每个像素点,颜色值大于均值  $\mu$  的赋值为 1,反之为 0,这样就得到了一系列大小为  $m \times m$  的二进制块。这些二进制块不仅体现了图像块内的纹理特征,而且在一定程度上反映了图像中的形状分布。相似的纹理结构会产生相同的纹理值,定义这些二进制块为图像的纹理基元。用与这些二进制块等值的十进制值来表示这些纹理基元的值。提取过程如图 5.24 所示 (这里  $m=2$ )。

图像块	<table><tr><td>20</td><td>22</td></tr><tr><td>8</td><td>7</td></tr></table>	20	22	8	7	<table><tr><td>9</td><td>19</td></tr><tr><td>11</td><td>20</td></tr></table>	9	19	11	20	<table><tr><td>17</td><td>11</td></tr><tr><td>18</td><td>9</td></tr></table>	17	11	18	9	<table><tr><td>20</td><td>7</td></tr><tr><td>6</td><td>23</td></tr></table>	20	7	6	23	<table><tr><td>8</td><td>7</td></tr><tr><td>6</td><td>8</td></tr></table>	8	7	6	8
20	22																								
8	7																								
9	19																								
11	20																								
17	11																								
18	9																								
20	7																								
6	23																								
8	7																								
6	8																								
二进制块	<table><tr><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td></tr></table>	1	1	0	0	<table><tr><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td></tr></table>	0	1	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td></tr></table>	1	0	1	0	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1
1	1																								
0	0																								
0	1																								
0	1																								
1	0																								
1	0																								
1	0																								
0	1																								
1	0																								
0	1																								
二进制码	1100	0101	1010	1001	1001																				
纹理值	12	5	10	9	9																				
	(a)	(b)	(c)	(d)	(e)																				

图 5.24 图像块与相应的纹理基元值

在提取图像的纹理基元时,会出现如图 5.24 中 (d) 和 (e) 所示的情况,不同结构的块可能会产生相同的纹理基元值。因此,在算法中,需设定一个阈值  $\beta$ ,当图像块的平均灰度差小于这个阈值时,就把这个块看做是均匀块,纹理基元值设为 0;大于这个阈值时,就按上述方法计算它的纹理基元值。在实验中,我们采用了统计的方法来确定  $\beta$  值,经过多次实验,发现当采用图像预处理中第一种方法时,这个阈值为整个图像平均颜色差的 0.0025 时将其看做均匀块,不影响算法的精度;当采用第二种方法时,这个阈值取整个图

像平均颜色差的 0.0175 时可以取得满意的效果。在提取了图像的纹理基元后,可采用纹理基元值作为整幅图像纹理特征的描述。如图 5.25 所示,图(a)为示例图像,图(b)为用纹理基元表示的示例图像。从图中可以看出,纹理基元在一定程度上可以很好地反映图像的纹理特性。在实际应用中,虽然对图像分块越多越能正确体现图像的特征,但是过细的分割除了造成信息量的急剧增加外,也会导致描述一般性的损失。经过大量实验,将图像划分为  $2 \times 2$  大小的块是合理的。

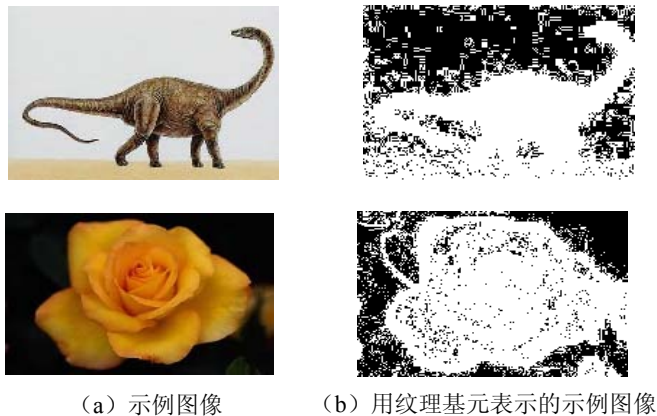


图 5.25 示例图像及相应的纹理图像

### 5.6.3 纹理基元共生矩阵

灰度共生矩阵所提取的图像纹理特征在区分不同的纹理上具有较好的效果。矩阵中的元素反映了图像灰度分布关于方向、局部邻域和变化幅度的综合信息,直接处理的基本单元对象是像素,但它提取的是整幅图像不同方向的纹理特征,并不能直接提供区别纹理的特性,还需要从矩阵中进一步提取有意义的统计量。而结构分析法是力求找出纹理基元,从结构上探索纹理规律,其基本思想是:复杂的纹理可以由一些简单的纹理基元按一定规律重复排列组合而成,然而缺乏纹理分布的空间信息。本小节利用定义的纹理基元,结合灰度共生矩阵,类似于灰度共生矩阵中的像素单元那样,研究其在空间上的分布规律,构造一种基于纹理基元的共生矩阵,然后从矩阵中提取有意义的统计量作为图像的纹理特征,从而将纹理分析中的结构法和统计法有机地结合起来。

设  $I'$  是  $I$  提取了纹理基元后  $M \times N$  大小的图像,图像中的纹理基元值为  $[0, n-1]$ 。则基于纹理基元的共生矩阵中的元素,即

$$H(i, j) = \eta\left((p(x, y), p(N_{(x, y)}))\right) = \alpha \sum_{x=1}^N \sum_{y=1}^M C_i(x, y) \sum_{(x', y') \in N(x, y)} C_j(x', y') \quad (5-156)$$

式中,  $i, j = 0, 1, 2, \dots, n-1$ ,  $p(x, y)$  是  $I'$  中  $(x, y)$  处的纹理基元值。 $N_{(x, y)}$  表示  $I'$  中  $(x, y)$  的四邻域,  $p(N_{(x, y)})$  则表示  $I'$  中  $(x, y)$  的四邻域处的纹理基元值。公式中  $C_i(x, y)$  的取值如下:

$$C_i(x, y) = \begin{cases} 1 & \text{如果 } p(x, y) = i \\ 0 & \text{其他} \end{cases} \quad (5-157)$$

于是,改进的共生矩阵中的元素  $H(i, j)$  就表示  $I'$  中所有  $(x, y)$  处的像素与其四邻域内的像素构成的像素对等于  $(i, j)$  的数目。考虑到图像的大小对矩阵的影响,于是对矩阵进行归一化



处理。因为基元对  $(p(x,y), p(N_{(x,y)}))$  在图像中大约为  $(m \times N \times M)$ ，所以设定  $\alpha$  取值为  $1/(m \times N \times M)$ 。

用该矩阵描述图像的纹理，避免了传统的共生矩阵只能描述单方向信息而带来的问题。该共生矩阵反映了图像灰度分布关于方向、局部领域和变化幅度的综合信息，但它并不能直接提供区别纹理的特征，还需从中进一步提取有用的统计量构成纹理特征。根据式 (5-12)，式 (5-13)，式 (5-14)，式 (5-16) 分别提取四个典型的统计量：能量、对比度、相关性和熵，来反映纹理不同方面的特性。由上述四个特征组成图像的纹理特征向量来分析及比较含有纹理结构的图像。

## 参 考 文 献

- [1] Alexandrov A.D., Ma W., Abbadi A., et al, 1995. Adaptive filtering and indexing for image databases, SPIE, 2420: 12-23.
- [2] 安志勇, 王晓华, 赵珊等. 一种图像纹理特征检索算法. 西安电子科技大学学报(自然科学版), 2007, 34 (3): 409-413.
- [3] 安志勇, 曾智勇, 赵珊等. 基于纹理特征的图像检索. 光电子激光, 2008, 19 (2): 230-232.
- [4] Ballard D. H., Brown C. M., 1982. Computer Vision. Prentice-Hall.
- [5] Beck J., 1966. Effect of orientation and of shape similarity on perceptual grouping. Perceptual psychophysics, 1 (7): 300-302.
- [6] Beddow J.K., 1997. Image analysis sourcebook. American Universities Science and Technology Press.
- [7] Bergen J.R., Julesz B., 1983. Parallel versus serial processing in rapid pattern discrimination. Natural, 303 (7): 696-698.
- [8] Bergen J.R., Adelson E.H., 1999. Early vision and texture perception using feature distribution. Pattern Recognition, 32 (3): 447-486.
- [9] Brochard J., Khoudeir M., Augereau B., 2001. Invariant feature extraction for 3D texture analysis using the autocorrelation function. Pattern Recognition, 22: 759-768.
- [10] Carlucci L., 1972. A formal system for texture languages. Pattern Recognition, 4 (1): 53-72.
- [11] Castleman K R 著. 朱志刚等译. 数字图像处理. 北京: 电子工业出版社, 2002.
- [12] Chang T., Jay Kuo C. C., 1993. Texture analysis and classification with true-structured wavelet transform. IEEE Trans. on Image Processing, 2 (4): 429-441.
- [13] Chaudhuri B.B., Sarkar N., 1995. Texture segmentation using fractal dimension. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 (1): 72-77.
- [14] Coggins J.M., Jain A. K., 1985. A spatial filtering approach to texture analysis. Pattern Recognition, 3: 195-203.
- [15] Dwong M.K., Lin B., 1995. W-transform method for feature-oriented multiresolution image retrieval. SPIE, 2941: 1086-1095.
- [16] Efros A.A., Leung T. K., 1999. Texture Synthesis by Non-parametric Sampling. IEEE International Conference on Computer Vision, 1033-1038.
- [17] Jain A.K., Farrokhnia F., 1991. Unsupervised texture segmentation using Gabor filters. PR, 24 (12): 1167-1186.

- [18] Jayaramamurthy S.N., 1980. Texture discrimination using digital de-convolution filters. Proc. SICPR, 1184-1186.
- [19] Julesz B., 1975. Experiments in the visual perception of texture. Scientific American, 232: 34-43.
- [20] Haralick R.M., Shanmugam K., 1973. Texture features for image classification. IEEE Trans. On system, man and cybernetics, SMC-3 (6): 610-621.
- [21] Hawkins J.K., 1970. Texture properties for pattern recognition. Picture Processing and Psychopictorics, New York: Academic Press, 347-370.
- [22] He Dong-chen, Wang li, 1990. Texture Unit, texture spectrum, and texture analysis. IEEE Transactions on Remote Sensing, 28 (4): 509-512.
- [23] 洪继光. 灰度-梯度共生矩阵纹理分析方法. 自动化学报, 1984, 10 (1): 22-25.
- [24] 黄祥林. 基于压缩域的图像检索技术的初步研究. 博士学位论文, 北京工业大学, 2001.
- [25] 黄元元. 基于视觉特征的图像检索技术研究. 博士学位论文, 南京理工大学, 2003.
- [26] Gagaudakis G., Rosin P., 2002. Incorporating Shape into Histograms for CBIR. Pattern Recognition, 35 (1): 81-91.
- [27] Galloway M.M., 1975. Texture analysis using Grey-level Run Lengths. Computer Graphics and Image Processing, 4: 172-179.
- [28] Geman S., Geman D., 1984. Stochastic relaxation Gibbs distribution, and the Bayesian restoration of images. IEEE Trans Pattern Anal Machine Intell, PAMI-16: 721-741.
- [29] Gerald Schaefer, Simon Lieutaud, 2004. Image retrieval based on block colour co-occurrence matrix and pattern histogram. International Conference on Image Processing, 413-416.
- [30] Gross M.H., Koch R., Lippert L., et al., 1995. Multiscale image texture analysis in wavelet space. In: Proc of IEEE Int Conf on Image Processing. Washington DC, 2 (4): 256-259.
- [31] Karkanis S., Galousi K., Maroulis D., 1999. Classification of Endoscopic Images based on texture spectrum. Workshop on machine learning in medical applications, 5-16.
- [32] Khotanzad A., Kashyap R.L., 1987. Feature selection for texture recognition based on image synthesis. IEEE Transactions on Systems, man and Cybernetics, 17 (6): 1087-1095.
- [33] Laine A., Fan J., 1993. Texture classification by wavelet packet signatures. IEEE Trans on Pattern Analysis and Machine Intelligence, 15 (11): 1186-1191.
- [34] Laws K. L., 1980. Rapid texture identification. SPIE, 238: 376-380.
- [35] 李永根, 吴纪桃. 分形与小波. 北京: 科学出版社, 2002.
- [36] Liu F., Picard R.W., 1994. Periodicity, directionality, and randomness: Wold features for perceptual pattern recognition. Proceeding of 12th International Conference on Pattern Recognition, 184-189.
- [37] Lowitz G.E., 1983. "Can a local histogram really map texture information?". Pattern Recognition, 16 (2): 141-147.
- [38] Lu S.Y., Fu K.S., 1978. A Syntactic Approach to Texture. Analysis. CGIP, 7: 303-330.
- [39] Luo J. Savakis A.E., 2001. Self-supervised texture segmentation using complementary types of features. Pattern Recognition, 34 (11): 2071-2082.
- [40] Manjunath B.S., Ma W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (8): 837-842.
- [41] Mao J., Jain A.K., 1992. Texture classification and segmentation using multiresolution simultaneous

- autoregressive models. *Pattern Recognition*, 25 (2): 173-188.
- [42] 孟祥增. 基于语义的 WEB 图像检索研究. 博士学位论文, 北京邮电大学, 2004.
- [43] Ojala T., Pietikainen M., Harwood D., 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29: 51-59.
- [44] Ojala T., Pietikainen M., Maenpaa T., 2002. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24: 971-987.
- [45] Ortega M., 1997. Supporting similarity queries in MARS. *Proc.ACM Multimedia*, 403-413.
- [46] Pentlind P., 1984. Fractal-based description of natural scenes. *IEEE Trans on PAMI*, 6 (6): 661-674.
- [47] Picard R.W., Kabir T., Liu F., 1993. Real-time recognition with the entire Brodatz texture database. New York: *Proc IEEE Int Conf on Computer Vision and Pattern Recognition*, 638-639.
- [48] Qiu guoping, 2003. Color image indexing using BTC. *IEEE Transactions on Image Processing*, 12 (1): 93-101.
- [49] Rama Chellappa, 1985. Two-dimensional discrete Gaussian Markov Random Field Models for image processing. *Progress in Pattern Recognition*, 2: 79-112.
- [50] Rao A.R., 1990. A taxonomy for texture description and identification. Springer-Verlag.
- [51] Remash N., Sethi I.K., 1995. Feature identification as an aid to content-based image retrieval. *SPIE*, 2420: 2-11.
- [52] 任仙怡, 张桂林, 陈朝阳. 基于纹理谱纹理分割方法. *中国图像图形学报*, 1998, 3 (12): 983-986.
- [53] 阮秋琦. 数字图像处理学. 北京: 电子工业出版社, 2001.
- [54] Russ J.C., 2002. The image processing handbook. 4<sup>th</sup> Ed. CRC Press.
- [55] Shapiro L., Stockman G., 2001. *Computer Vision*. Prentice Hall.
- [56] 盛文, 杨江平, 柳建等. 一种基于纹理元灰度模式统计的图像纹理分析方法. *电子学报*, 2000, 28 (4): 74-77.
- [57] Stack H.G., Laufkotter G.P., 1995. Image indexing and content-based access to databases of medical image with wavelet. *SPIE*, 2569: 790-799.
- [58] 孙即祥, 王晓华. 模式识别中的特征提取与计算机视觉不变量. 北京: 北京工业出版社, 2001.
- [59] Takahashi N., Lwasaki M., Kunieda T., et al, 2000. Image retrieval using spatial intensity features. *Signal Processing: Image Communication*, 16: 45-57.
- [60] Tamura H., Mori S., Yamawaki T., 1978. Texture feature corresponding to visual perception. *IEEE-SMC*, 8 (6): 460-473.
- [61] Thyagarajan K., Nguyen J., Persons C., 1994. A maximum likelihood approach to texture classification using wavelet transform. *Proc IEEE Int Conf on Image Processing*, 640-644.
- [62] Tuceryan M., Jain A. K., 1990. Texture segmentation using Voronoi Polygons. *IEEE Trans on PAMI*, 12: 211-216.
- [63] 王波, 姚宏宇, 李弼程. 一种有效的基于灰度共生矩阵的图像检索方法. *武汉大学学报信息科学版*, 2006, 31 (9): 761-764.
- [64] 王海霞. 基于纹理特征的图像检索技术研究, 硕士学位论文, 燕山大学, 2006.
- [65] 王亮申, 欧宗瑛. 图像纹理分析的灰度-基元共生矩阵法. *计算机工程*, 2004, 30 (23): 19-21.
- [66] 王文惠. 基于内容的图像检索技术研究. 博士学位论文, 国防科学技术大学, 2001.

- [67] Wu C.M., Chen Y.C., Hsieh K.S., 1992. Texture features for classification of ultrasonic liver images. IEEE\_MI, 11: 141-152.
- [68] 吴高洪, 章毓晋, 林行刚. 基于分形的自然纹理自相关描述和分类. 清华大学学报, 2000, 40 (3): 90-93.
- [69] 吴均, 赵忠明. 利用基于小波的尺度共生矩阵进行纹理分析. 遥感学报, 2001, 5 (2): 100-103.
- [70] 杨育彬. 基于内容的图像检索关键技术及其应用研究. 博士学位论文, 南京大学, 2003.
- [71] 章毓晋. 图像工程 (第2版). 北京: 清华大学出版社, 2007.
- [72] 赵珊, 孙君顶, 周利华. 基于方块编码的图像纹理特征提取及检索算法. 光电子激光. 2006, 17 (8): 1014-1017.
- [73] 赵珊, 孙君顶, 安志勇等. 一种新的基于纹理基元的图像检索算法. 第一届人机和谐环境学术会议论文集. 云南, 2005: 394-399.
- [74] 赵珊, 孙君顶, 周利华. 一种新的基于关键子块的图像检索算法. 光子学报, 2007a, 36 (2): 376-379.
- [75] 赵珊, 安志勇, 周利华. 基于方块编码的图像特征提取及检索算法. 光电工程, 2007b, 34 (1): 117-120.
- [76] 周明全, 耿国华, 韦娜. 基于内容图像检索技术. 北京: 清华大学出版社, 2007.

## 第6章 基于压缩域的图像检索

随着许多压缩标准（如 JPEG、MPEG 和 H.261 等）的制定和推广，压缩图像的使用越来越普遍和广泛。直接对压缩图像进行检索，即压缩域上的检索技术，也得到了广泛关注和研究。区别于传统的基于原始数据域的图像检索，压缩域的图像检索技术是通过挖掘图像压缩时的中间结果或最终码流中包含的信息，力争在不解码或部分解码的情况下提取表征图像内容的特征，并以此作为索引实现基于内容的图像检索，是直接面向数据量大大减少（但信息量没有或基本没有减少）的压缩数据进行的检索。本章从常用的几类图像压缩算法及静止图像压缩标准入手，系统介绍了近年来国内外学者的一些研究成果，详细论述了 DCT 压缩域中两种特征提取算法。

### 6.1 概述

前面讨论的图像特征，主要考虑的是使用直接采集的原始格式的图像，而对于普遍存在的压缩形式的图像数据，须先解压缩才能进行检索用特征的提取。对原始数据的分析处理，意味着其处理数据量要高于压缩格式数据量的几倍、十几倍甚至几十倍，不但计算量大，而且需占用较多的中介存储空间，极不利于在计算资源有限的环境里进行，无疑将影响图像检索系统的实用性和灵活性。特别是对一些大型的图像检索系统和要求实时性的动态检索系统，这种传统的处理模式往往难以满足要求。在这样的背景下，压缩域图像检索技术的研究受到了广泛的重视。在传统的图像检索技术中，操作对象是原始数据（可以直接获取的或者是对压缩数据进行解码以后得到的），而在基于压缩域的图像检索中，则试图直接在压缩数据上进行操作，不需要或不完全需要解压缩的环节。事实上，由于传统图像检索技术与压缩技术相互孤立，现有的压缩编码算法所形成的压缩码流只考虑了存储和传输的需要，即在可视质量的允许下尽一切可能用最少的比特数来表征图像信息，而未考虑后续分析处理的需要，因此并不具有支持图像检索的能力。而压缩域图像检索技术则从二者结合的角度进行检索技术的研究，实际上是把图像的压缩技术与检索技术融合在一起，克服了现有的将压缩技术与检索处理技术相分离所带来的局限性，力图实现快速、高效、灵活的检索算法。基于压缩域的图像检索技术与传统的基于原始域或解压域的图像检索技术相比有许多优点，包括（邵虹 2004）：

- （1）在压缩域上检索可省略解压缩的附加环节，既可减少处理时间，也可减少设备开销；
- （2）许多图像压缩算法已对图像进行了大量的处理和分析，在检索中利用这些处理和分析的结果，可减少计算量，提高检索效率；
- （3）压缩域上的数据量比原始域或解压域上的数据量少，这有利于提高整个系统的效率，尤其是在检索系统要求实时响应的场合；
- （4）基于特征的图像检索方法在存储图像时（建库）除了存储图像外还要存储相应的

特征向量,而在压缩域上,某些特征向量的信息已经包含在压缩系数中,所以额外的存储量可以省去。

基于压缩域的图像检索技术处理速度快,实时性好,计算量小的特性,如果能将压缩域提取的某些特征用 MPEG-7 所规定的特征描述符表示,则其检索技术更具有广泛意义;另外,由于对计算资源的需求减少,基于压缩域的图像检索技术非常有利于手持网络终端等计算资源有限的环境,对网络信息安全也具有重要的作用;而且压缩域的处理技术可以在不解压的情况下快速、高效地监测多媒体信息的安全性。这些,将为我国的信息安全提供新的技术基础。因此基于压缩域的图像检索技术具有非常广阔的应用前景,对多媒体业务的普及、拓展和我国信息安全起着积极的作用。目前,基于压缩域的图像检索技术的研究已经引起了国内外的广泛关注,成为一个国际性研究热点(赵珊 2007a)。

### 6.1.1 图像压缩技术

图像数据的高效压缩和有效检索是构建多媒体信息业务和应用所必须解决的两个关键问题,对压缩域的图像进行检索必须根据其所采用的压缩技术而选择合适的检索技术。对图像压缩来说,任何压缩机制的基本思想都是去除数据中存在的冗余性。所谓冗余性,就是信息在一定程度上的重复,可以根据给出的一部分数据来判断其相邻的数据。图像的压缩机制可以分为两种,有损压缩和无损压缩。

(1) 无损压缩。即冗余度压缩方法,也称信息保持编码或熵编码,利用数据的统计特征进行压缩,解码图像和压缩编码图像严格相同,没有任何失真和信息损失。从数学上讲是一种可逆运算,但压缩率受到数据统计冗余度的理论限制,压缩率一般为  $2:1 \sim 5:1$ ,广泛用于文本数据、程序和特殊应用场合的图像数据(如指纹图像、医学图像等)的压缩。由于压缩比的限制,仅使用无损压缩方法不可能解决图像和视频的存储和传输问题。

(2) 有损压缩。即信息量压缩方法,也称失真度编码或熵压缩编码,利用人类视觉对图像中的某些频率成分不敏感的特性,允许压缩过程中损失一定的信息,但是损失的部分对理解原始图像的影响较小,从而以一定的失真换来大的高的压缩率。有损压缩广泛应用于语音、图像和视频数据的压缩。

图像压缩的主要目标就是在给定目标比特率的情况下,使图像的质量损失最小,或者是在给定图像质量的情况下,使图像表征所需要的比特数量最少,从而有效降低图像存储空间或减少图像传输的带宽。在过去的几十年中已经发展了许多图像压缩编码方法,这些方法从不同的途径消除图像中的冗余,以达到压缩图像数据的目的。

行程编码(Run Length Coding, RLC)是一种相对简单的编码方案,是指在每一行扫描的像素中比较相邻像素的幅度(如高度等),当幅度有一个显著变化时,就说有一个行程存在,像素的连续长度和终点位置标记是其重要的参数。根据终点位置标记方法的不同,行程编码可以分为两类:行程终点编码和行程长度编码。前者的终点编码位置由扫描行的起始点算起至行程终点位置的像素数确定,而对后者来说,某行程长度的终点位置由它距前一终点的距离位置来确定。行程终点编码又分为两种:①线性码法,根据不同行程长度赋予不同的码字,大行程码字长,小行程码字短;②对数码法,它的码字长与行程长度的对数成正比。行程编码的码字结构相对较简单,适用于二值图像(如传真)的编码,因为1和0总是交替出现,对于不同串长度按其发生概率不同分配以不同码字,可以将“1”值的长度和“0”值的长度单独编码,也可将二者长度混合编码。行程编码的优点是非常适于

大面积、相同码值较多的情况；缺点是对误码很敏感，一个值传输或存储出错，可能导致整个图像混乱，而且对不连续的情况编码效果很差，甚至可能出现编码后的代码数比编码前还多。

霍夫曼编码 (Huffman Coding) 是一种较有效的方法，是一种长度不均匀，平均码率可以接近信息源熵值的一种编码。它的基本编码思想是对于出现概率大的信息采用短字长的码，对于出现概率小的信号采用长字长的码，以达到缩短平均码长，从而实现数据压缩的目的。霍夫曼编码的最高压缩效率可达到 8:1，但是在一般实施过程中，很难达到这种压缩比，若图像中存在某个拥有长行程的字节值时，使用行程编码压缩方式可能会更好 (Wu D. et al 1999)。而近几年发展起来的 Rice 编码是一种特殊的霍夫曼编码方法，它在编码时不需要码表，但却能提供与使用多个霍夫曼码表相同的功能，Rice 编码的速度优于广泛采用的算术编码，且具有接近于算术编码的效率，因而该编码方法在无失真图像压缩领域引起了重视 (赵德斌等 1998)。目前游程长度编码、霍夫曼编码和算术编码已被 JPEG、MPEG 等压缩标准所普遍采纳，用于对变换、预测编码之后的图像系数在无失真的前提下进一步提高压缩比。

字典表压缩方法 (Lemple-Ziv&Welch, LZW) 也是一种有效的无损压缩方法。目前广泛采用的 LZW 压缩方法有两种类型。一种类型是在数据压缩过程中，寻找当前等待进行压缩处理的数据串是否在已经处理的数据串中出现过，如果曾经出现过，则利用指向该数据串的指针代替当前进行压缩的数据串，此时，字典是隐式的，它用曾经处理过的数据描述。另一种类型是为输入数据创建一个短语字典，如果在当前等待进行压缩的数据流中发现在字典中已经存在相应的短语，则利用该短语的相应索引取代原始数据。LZW 方法的特点是压缩效率很高，但比较复杂，不仅用于文字数据的压缩，还可以成功地用于某些图像的压缩处理。

算术压缩方法 (Arithmetic Coding, ARC) 与霍夫曼编码压缩方法相似，都是利用比较短的代码取代图像数据中出现比较频繁的数据，而利用比较长的代码取代使用频率比较低的数据，从而达到数据压缩的目的。它同时又采用了字典压缩编码的思想，不仅压缩数据值，而且压缩值序列，从而达到更理想的压缩比例，尤其适合大多数数据由相同的重复序列组成的图像文件。其基本思想是：将每个不同的序列按照出现的频率映射到 0 和 1 之间的相应数字区域内，该区域表示成可以改变精度的二进制小数，其中出现频率越高的数据利用精度越高的小数表示。算术编码算法可以大幅度地减少文件长度，甚至可以达到 100:1 的压缩，在 JBIG 与 JPEG 图像文件格式的数据处理步骤中占据很重要的地位。由于压缩算法比较复杂，其同时还受到几项 IBM 的专利保护，导致了算法许可性的不确定性 (张问银 2005)，阻碍了算术压缩算法的推广。针对不同的图像文件，算术压缩的压缩比主要与源文件的数据分布及其标准模式的精度有关。

预测编码 (Predictive Coding) 是研究最早的图像压缩方法，它的基本思想是：利用图像数据所具有的空间或时间相关性，用相邻的已知像素 (或图像块) 来预测当前像素 (或图像块) 的取值，然后对预测误差进行量化和编码。预测编码的关键在于预测算法的选取，这与图像信号的概率分布有很大的关系，实际应用中常根据大量的统计结果采用简化的概率分布形式来设计最佳的预测器，有时还使用自适应预测器来刻画图像信号的局部特性，以提高预测效率。预测编码方法计算简单，但由于是基于差值信号的统计特性发展起来的，故存在着一些缺点，例如，对黑白灰度有突变的点，会有较大的预测误差，使重建图像的

边缘模糊,分辨率降低;对图像亮度值变化缓慢的区域,其差值信号为零,但因预测值偏大而使重建图像产生噪声。

变换编码是消除图像数据空间相关性的一种更为有效的图像压缩方法。它通过正交变换实现数据压缩,图像经过正交变换后,再经过多维坐标系中适当的坐标旋转和变换,能够把散布在各个坐标轴上的原始图像数据,在新的适当的坐标系中集中到少数坐标轴上,因而只需对少数的高能量系数进行适当的量化和熵编码,就可以用较少的编码比特数来表示一幅子图像,实现图像的压缩编码。图像变换编码从数学的角度看,可用于图像变换编码的正交变换除 Fourier 变换, Walsh-Hadamard 变换外,还有离散余弦变换(Discrete Cosine Transform, DCT)、哈尔变换、K-L 变换等,不同的变换有不同的压缩效果(压缩比和重建图像质量)。目前最流行的静态图像压缩标准 JPEG 采用的是离散余弦变换,然而基于 DCT 的变换编码的主要不足在于变换方法不具有好的时频局域性和全局变换,因而在提高编码压缩比时会出现块效应,导致图像的边缘轮廓模糊,严重影响重建图像的主观质量。

矢量量化(Vector Quantization, VQ)是另一种高效的图像压缩技术。根据 Shannon 信息率失真理论,只要码书数足够大,矢量维数足够大,矢量量化方法能够得到 Shannon 定理所规定的下限(沈兰荪 1998)。实际的矢量量化中编码端和解码端具有相同的码书(Codebook),码书由所有可能矢量值集合的有序子集组成,具体编码时,编码端依据选定的距离测度(或称代价函数)在码书中对输入的图像分块矢量进行匹配,全面质量管理对匹配码的码字序号进行编码,从而实现由一个矢量所需比特数到一个码字序号所需比特数的压缩。矢量量化方法实现的关键在于最佳码书的设计和快速匹配算法的选择,码书性能的好坏直接影响着整体矢量量化的性能。传统的码书设计算法,如 C-均值算法和 LBG 算法,简单直观且收敛速度快,但它们的聚类结果过于依赖初始训练矢量的选择,极易陷入局部最小。近年来,许多学者利用一些新兴理论,如神经网络和模糊理论等,对码书的设计算法进行了深入的研究,提出一些比传统码书设计算法性能更优的算法,如竞争性学习算法(Ahalt S.C. et al 1990)、模糊竞争性学习算法(Chung F.L. et al 1994)等,其中竞争性学习算法可以说是传统 LBG 算法的自适应形式,因而它仍保留 LBG 算法的一些缺陷,如存在码矢欠利用问题。为此人们进一步将模糊理论引入到竞争性学习算法中,提出了模糊竞争性学习算法。与传统神经网络竞争性学习算法不同,模糊竞争性学习算法中每个权矢量都按照隶属度函数得到不同程度的调整,较充分地利用了输入训练矢量与码矢之间的距离信息来修正模糊隶属度函数,使之有效地控制码矢的训练调整过程,从而较好地解决码矢欠利用等问题。

金字塔编码方案是把图像分解成许多不同分辨率的子图像,并把高分辨率(尺寸较大的)的子图像放在下层,把低分辨率(尺寸较小的)的子图像放在上层,从而像一个金字塔,借助拉普拉斯金字塔算法,对图像的每一层分别量化、编码,并对视觉不敏感的层进行粗化,用较少的码字编码,从而达到压缩的目的。解码是逐层累积,从轮廓到细节重建图像。这样的等级图像结构特别适于检索型的应用场合,并且可以根据需要给不同的塔层赋予不同的优先级。图像的金字塔表示方法可以用很紧凑的方法来进行图像的编码,但它的缺点在于编码过程中会增加一定的数据量,因而无法达到高的压缩比。

方块编码(Block Truncation Coding)的基本思想是:将图像分割成相互不覆盖的  $n \times n$  的像素块,由于子块内各邻近像素间具有亮度互相近似的相关性,可选用两个适当的亮度级  $(a_0, a_1)$  来近似代表子块内各像素原来的亮度。通常可利用均方误差最小的方法,来逐个



求出各子块的这两个有代表性的亮度级。其编码原理在第5章已介绍过,这里不再重复。

以上的几种压缩方法在20世纪80年代中期以前的研究中占据着重要的地位,而80年代中后期,相关学科的迅速发展和新兴学科的不断涌现为图像压缩编码的研究发展注入了新的活力。许多学者结合模式识别、计算机图形学、计算机视觉、神经网络、小波变换和分形几何等理论对图像压缩编码的新途径进行了探索。同时关于人类视觉生理、心理特性的研究成果也为图像压缩提供了新的思路,许多新的方法,如第二代图像压缩编码方法、基于分形的图像压缩、基于模型的图像压缩方法以及基于小波变换的图像压缩算法相继被提出来(沈兰荪 1998)。

子带编码是第一代编码方法和第二代编码方法的过渡。它的基本思想是:把图像信号通过一组带通滤波器分解成不同频带内的分量,然后在每个独立的子带中,对信号进行降速采样和单独编码。首先将图像进行滤波而产生一个图像集合,其中每个都占据一个空间频率的限制范围,这些图像就称为子带。由于子带与原来满频带图像相比,均占有一个小的带宽,就可对它们进行降速采样,这种滤波处理和亚抽样称为图像的分解阶段。之后利用一个或多个编码器对子带进行编码,各子带的图像的统计特性是以人眼对它的敏感程度来确定的,可以根据信息论和人眼的视觉特性,利用不同的比特率甚至是不同的编码技术对各子带编码。对于子带编码序列的译码,则要使用升速取样以及适当滤波而获得重建的子带图像。最后将各子带相加,得到完整的重建图像。这一功能的实现环节称为合成阶段。在整个编码过程中,构成子带并不带来任何压缩效果,这是因为全部子带所需的样本数与图像所需的样本数相同。子带编码方式的积极意义在于,这些子带可以较原图像有更高的编码效率,即子带编码的高效不在于形成子带,而在于各子带的高效编码。子带编码的优点是,由于量化在各子带内单独进行,量化噪声便被限制在各子带内,可以防止能量较小的频带内信号受其他频带内量化噪声的干扰和影响。

第二代图像压缩方法又称结构编码,它的主要特点在于不仅局限于信息论的基础,还要求充分利用人眼的生理和心理特性以及信源的各种性质以期获得高压缩比。按处理方法不同可以分为基于分裂合并的方法和基于各向异性滤波器的方法。基于分裂合并方法先将图像分为纹理和边缘轮廓,然后各自采用不同的方法进行编码;而基于各向异性滤波器的方法,则是先对图像进行方向性滤波,得到不同方向的图像信息,再根据人眼的方向敏感性对各个通道采用特定的方法各自编码。

对于各种变换编码方法来说,都是先经变换,然后再对变换系数进行取舍、量化和编码以达到数据压缩的目的。其变换本身没有数据压缩作用,此外,变换方法大半要涉及较复杂的运算。神经网络的压缩方法正是基于以上两点对变换编码进行的有益改进。神经网络是一种模仿及延伸人脑的新型信息处理系统,是由大量全局、局部或稀疏连接的简单处理器组成的非线性动力学系统。它具有自学习特性、大规模并行处理、非线性处理及分布式存储四种特性,目前在图像编码中已获得成功的应用。基于神经网络的压缩技术采用两种方法发展了一种广义变换,即用Hopfield网络实时求解方法和前向网络方法。其优点是:一旦训练完成,压缩编码的计算就十分简单;压缩主要靠学习完成,编码可以进一步提高压缩比;对信号统计特性没有要求,因此适合于任何图像的压缩,且压缩比和信噪比可事先设置。其缺点是学习训练时间往往很长,且Hopfield网还有取局部极小值的问题。

分形图像压缩(Fractal Coding)是在分形几何理论的基础上发展起来的一种编码方法。分形理论是传统欧氏几何相关理论的扩展,它是研究不规则图形和混沌运动的一门非线性

科学,它描述了自然界物体所具有的自相似性。Barnsley M.最早引入了迭代函数系统来刻画图像中的自相似性,并将其用于图像压缩,对某些特定图像获得了极高的压缩比,但是他的方法需要人工干预。为此他的学生 Jacquin 提出了基于分块迭代函数系统的自动分形图像压缩方法,从而将分形在图像编码上的应用推进了一大步(Jacquin A.E. 1993)。分形图像压缩的关键在于图像 IFS (Iterative Function System) 码的获取,目前已有专利算法,对某些图像可获得 30~70 倍的压缩比。然而分形图像编码的理论基础一方面决定了它只有对那些明显具有自相似性或统计自相似性的图像才有较高的压缩率,对一般的图像,特别是相似性不强的图像,基于分形的方法其效率并不高;另一方面,分形压缩是一种非对称方法(编码复杂度远远高于解码复杂度),其实现的结构很难利用人眼的视觉特性,只有和其他编码方法相结合,才能得到较好的编码效率和较低的实现复杂度(Wohlberg B. et al 1999)。

模型基图像压缩方法(Model-based Coding)是一种被认为很有前景的低比特率编码方法,它利用了计算机视觉和计算机图形学中的方法和理论,在编码、解码两端分别建立了相同的模型,在编码端对输入图像进行分析,获取模型参数,并将模型参数传递给解码端;解码端根据接收到的模型参数进行图像合成,重建原始图像。因此模型基压缩方法的核心在于模型的建立和模型参数的获取(魏海 2000)。目前最受关注的模型编码方法的建模对象是图像中的人脸。根据已知的二维或三维人脸模型,获得简单的动画人脸只需 500~1000b/s 的开销,而获得较为复杂的动画人脸也只不过需要 5~10kb/s 的开销。但是基于人脸模型的编码方法要真正达到实用化水平,还有许多问题有待解决。例如,计算复杂度太高;求取建模对象的运动参数耗时太多;可以编码的图像场景单一、受限;在实际应用中极不灵活,情况稍有变化,模型就会失效。可以预期,随着计算机视觉研究的进一步进展,模型基编码方法将会具有更普遍的意义和较高的实用价值。

小波图像压缩方法是近年来受到广泛重视并已逐步得到应用的变换编码技术。与传统的 DCT 变换相比,离散小波变换(DWT)具有良好的时频局部化特性和较好的能量集中特性,同时由于小波变换是对整个图像进行的,因此,基于小波的图像压缩方法能够很好地克服方块效应;此外,小波变换的多分辨率特性提供了利用人眼各种视觉特性的良好机制。正是由于这些原因,众多学者对基于小波的图像压缩方法进行了研究。Shapiro J.M.利用小波分解不同频带之间的关联,采用访问演出树实现了自嵌套的图像压缩方法,在重建图像质量良好的条件下,获得了 128 倍的高压缩比(Shapiro M. 1993)。而零树编码所获得的高性能使得研究者们对其进行了许多改进,当前基于小波零树思想的小波图像压缩方法仍被认为是利用小波变换对图像进行压缩编码所获得的最好结果。因此,新一代的多媒体压缩标准 MPEG-4 中静态纹理的编码就是采用了基于零树思想的小波编码方法,而新一代静止图像的压缩标准 JPEG2000 也全面采用基于小波的压缩编码算法。这些都充分表明基于小波的图像压缩算法获得了极大的成功。然而,就目前小波压缩方法所获得的压缩比来看,基于小波的图像压缩方法所具有潜能还远远未发挥出来,在最佳小波基函数的选择、人眼视觉特性的利用、变换系数的有效组织以及与其他编码方法的有效结合等方面仍需进一步深入地研究。

### 6.1.2 静态图像压缩标准

自 Oliver 提出电视信号的线性 PCM 编码理论以来,图像压缩编码经历了半个多世纪的发展,已进入了广泛应用和深入研究的高速发展时期。目前,有代表性的静止图像国际

压缩标准主要有以下几种。

### (1) JBIG 图像无损压缩标准

JBIG (Joint Bi-level Image Experts Group) 是由 ISO/IEC 和 ITU (原 CCITT) 联合委员会于 1992 年提出和开发的适用于二值和低精度灰度 (低于 6bit/像素) 图像的无损压缩标准 (Arps R.B. et al 1994)。采用 2D 常规或自适应预测模式、自适应算术编码等方法, 能自适应图像的特征, 对不同的图像类型均具有“稳健性”, 因而具有较高的编码效率。它能满足标准化的要求, 并能改善图像质量与大小之间的矛盾。对于那些满足现在二值图像质量的用户, JBIG 能通过减小文件大小来节约巨大的费用开支, 能更快地传输图像文件, 能在相同的文件大小条件下提供更好的图像质量。由于在获取灰度图像时, JBIG 结构并不仅限于标准的 8bit, 因此, JBIG 能提供较小的灰度级文件; 又因为 JBIG 将灰度数据存储于独立的位平面中, 所以, 能高效地被应用任何在比特数的灰度数据上。

### (2) JPEG

JPEG (Joint Photographic Experts Group) 是联合图像专家小组的简称, 即国际电报电话咨询委员会 (CCITT) 和国际标准化协会 (ISO) 联合成立的一个图像专家小组。1991 年该小组提出 ISO CD10918 标准建议草案, 1992 年成为国际标准 ISO/IEC IS10918, 后来将该标准称为 JPEG。它主要涉及连续色调 (灰度和彩色) 静止图像的压缩编码。JPEG 标准支持两种图像模式: 顺序型和渐进型, 用以满足用户对具体应用的不同需要。JPEG 压缩算法分为两大类: 无失真压缩和有失真压缩。使用无失真压缩算法将源图像数据转变为压缩数据, 该压缩数据经过对应的解压缩算法处理后可以获得与源图像完全一致的重建图像; 有失真压缩算法主要基于离散余弦变换, 所生成的压缩图像数据经过解压缩生成的重建图像与源图像在视觉上保持基本一致, 压缩比越大, 视觉一致性越差。JPEG 共有四种编码模式: 顺序型、渐进型、无失真编码和分层编码。根据编码种类的不同将 JPEG 系统分成基本系统和扩展系统: 基本系统由 DCT 顺序模型及 Huffman 编码组成, 所有符合 JPEG 标准的设备都具备基本系统; 扩展系统提供不同的选项, 即除基本系统外的其他编码模式, 如渐进型编码、算术编码、无失真编码、分层编码等。

基于 DCT 变换的 JPEG 基本系统利用人的视觉特性, 通过 DCT 系数量化和无损压缩编码去掉视觉不敏感信息, 并最终实现图像数据压缩的目的。JPEG 压缩流程图如图 6.1 所示, 包括图像数据输入、基于 DCT 的编码和压缩图像数据流处理的输出三个步骤, 解码的过程沿着箭头相反的方向进行。DCT 编码器包括正向 DCT 变换器、量化器和熵编码器三个部分, 并且附有量化表的熵编码表 (哈夫曼表)。

#### ① 颜色空间转换

JPEG 算法与色彩空间无关, 编码源图像的输入, 可以是单色图像的灰度值, 也可以是彩色图像的亮度分量或色差分量信号。因此, 可以压缩来自不同颜色空间的数据, 如 RGB、YCbCr 和 CMYK。图像输入端, 首先将 RGB 空间转换到 YUV 空间, 亮度信号用 Y 保存, 色彩分量用两个颜色差值分量 U 和 V 来保存。

#### ② DCT 变换

在对图像进行 DCT 变换时, 考虑到编码效率和运算复杂度, 采用了分块技术, 一般采用  $8 \times 8$  分块, 因此首先对图像进行  $8 \times 8$  分块, 再分别对每块进行 DCT 变换。采样精度为 P 位, 把  $[0, 2^P - 1]$  范围内的无符号整数延拓为  $[2^{P-1} - 1, 2^P - 1]$  的有符号整数, 以此作为离散余弦正变换 (FDCT) 的输入, 即

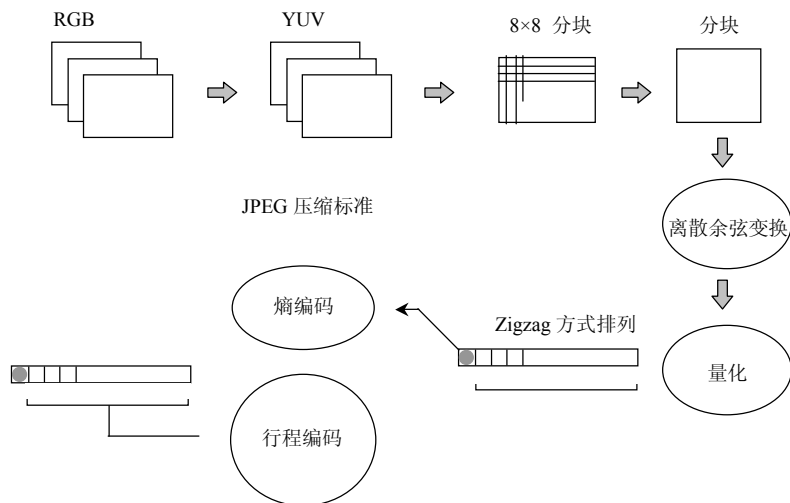


图 6.1 JPEG 压缩流程图 (赵珊 2007)

$$F(u, v) = \frac{1}{4} C_u C_v \sum_{i=0}^7 \sum_{j=0}^7 \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} f(i, j) \quad (6-1)$$

在解码器的输出端，离散余弦逆变换 (IDCT) 输出  $8 \times 8$  的数据块用以重构图像。IDCT 表示为

$$f(i, j) = \frac{1}{4} \left[ \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right] \quad (6-2)$$

图像经 DCT 变换后，输出 64 个 DCT 系数，对应于  $u=0, v=0$  的系数，称为直流分量，即 DC 系数，其余 63 个系数称为 AC 系数，即交流分量。DCT 变换实际上是像素域的低通滤波器，其低频分量都集中在左上角，且包含了图像的色彩、亮度等主要信息，高频分量则分布于右下角，代表图像的细节部分。图 6.2 给出了一幅  $8 \times 8$  的子图像块经过 DCT 变换后得到的变换系数。因为在一幅图像中像素之间的灰度和色差信号变化缓慢，在  $8 \times 8$  子块中像素之间相关性很强，所以经过 FDCT 变换后，在低频集中了数值大的系数，远离直流系数的高频交流系数则大多为零或趋于零，这就为数据压缩提供了可能。

### ③ 量化

DCT 变换后，其能量主要集中在变换域的左上角，即频率系数较低的区域，此时对变换域的系数值进行量化后，许多高频区域已经为零，特别有利于图像的压缩。这里的量化是根据人眼对频率的反映特性来设置的，其目的是尽量去掉引起视觉冗余的数据并保证图像的视觉质量。在 JPEG 标准中采用线性均匀量化器，对 64 个 DCT 变换系数  $F(u, v)$  除以量化步长，四舍五入取整。定义为

$$F^Q(u, v) = \text{Round} \left( \frac{F(u, v)}{Q(u, v)} \right) \quad (6-3)$$

式中， $Q(u, v)$  是量化步长，是量化表的元素，取值为 1 到 255 之间的任意整数。量化表元素随 DCT 变换系数的位置而改变，与 64 个变换系数一一对应，这是一个多到一的映射，也是 DCT 编解码信息损失的根源。JPEG 压缩标准中采用的两个量化表实例如图 6.3 所示。

139	144	14	153	155	155	155	155
144	151	15	156	159	156	156	156
150	155	16	163	158	156	156	156
159	161	16	160	160	159	159	159
159	160	16	162	162	155	155	155
161	161	16	161	160	157	157	157
162	162	16	163	162	157	157	157
162	162	16	161	163	158	158	158

子图像块

235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	13
-226	-18.5	-6.2	-3.2	-2.9	-0.1	-0.4	-1.2
-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	13
1.8	-0.2	-1.6	-0.3	-0.8	1.5	1.0	-1.0
-1.3	0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

DCT 变换系数

图 6.2 图像块及 DCT 变换后的 DC 系数及 AC 系数

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

亮度量化表

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

色度量化表

图 6.3 JPEG 压缩标准中采用的两个量化表实例

在接收端要进行逆量化，逆量化的计算公式为

$$F^{Q'}(u,v) = F^Q(u,v)Q(u,v) \quad (6-4)$$

#### ④ 编码

对量化后的数据，即 DC 系数和 AC 系数分别进行编码。DC 系数是  $8 \times 8$  子块 64 个采样的平均值。因为相邻的  $8 \times 8$  子块的相关性强，相邻块的 DC 系数差值很小，采用差分编码 (DPCM) 的方法，对  $\Delta DC_i = DC_i - DC_{i-1}$  进行编码可以用较少的位数，提高压缩比。经量化后的 63 个 AC 系数，根据其系数的频率分布特点进行处理。由于低频部分反映了图像的主要内容，通常具有较大的系数值，而高频部分系数较小，量化后 0 值较多，对所有的频率系数按 Zigzag 方式重新排列 (如图 6.4 所示)，使系数按照递增的空间频率定性地排列，保证低频分量先出现，高频分量后出现，会出现成片的零值，这时采用 RLE 编码的方式能有效的压缩成片的零值区域。

经过行程编码后，JPEG 标准对得到的数据再进行熵编码，旨在从统计的角度进一步降低数据量。使用熵编码可以对 DPCM 编码后的直流 DC 系数和行程编码后的交流 AC 系数做进一步的压缩。JPEG 建议两种熵编码方法：哈夫曼编码 (Huffman) 和自适应二进制算法编码 (Adaptive Binary Arithmetic Coding)。熵编码的过程大致可以看成由两个步骤组成，

首先将重新排序并且经过行程长度编码之后的量化系数转换成为符号队列，然后将符号队列转换成数据流。符号的定义与之前的 DCT 变化以及之后的熵编码操作无关，仅代表相应的数据，以供进行统计编码使用。熵编码从统计的角度进一步减少数据的冗余度，去掉了一些对视觉不重要的信息，减少了数据冗余，从而达到一定的压缩目的。

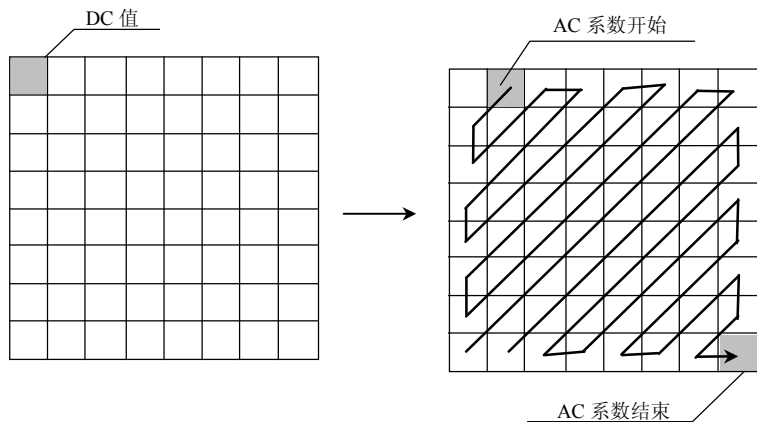


图 6.4 变换后的 DC 系数和 AC 系数以及 AC 系数编码采用的排列顺序

### (3) JPEG2000

JPEG2000 标准与传统的 JPEG 标准最大的不同，在于它扬弃了以离散余弦变换为主的块编码方法，而改采用以小波变换为主的多解析编码方式。作为 JPEG 标准的一个更新换代标准，它的目标是进一步改进目前压缩算法的性能，以适应低带宽、高噪声的环境，以及医疗图像、电子图书馆、传真、Internet 网上服务和保安等方面的应用。JPEG2000 规定了一系列对连续色调、二值、灰度或彩色数字静止图像的无失真或有失真编解码方法，JPEG2000 基本编码框图如图 6.5 所示。

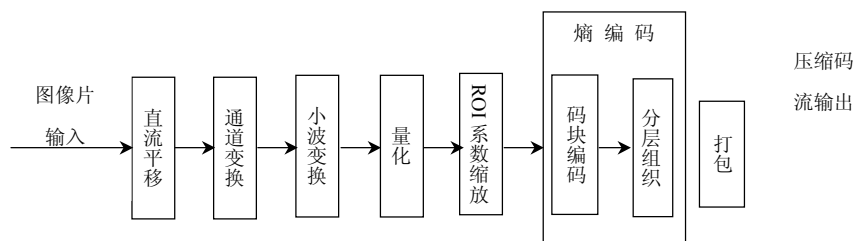


图 6.5 JPEG2000 基本编码框图（李晓华 2001）

在 JPEG2000 标准中，一幅图像首先被分解为若干方形的图像片，图像片是图像编解码的基本单位，各个图像片的编解码过程是独立进行的，分片的目的有两个，降低内存要求，实现特定区域的编解码。

① 直流模块是可选模块，仅适用于图像分量的采样值为无符号数的情况，其目的是去掉图像片中的直流分量，从而使小波变换后系数的取正值和取负值的概率基本相等，提高后续的自适应熵编码效率。

② 分量变换模块也是可选模块，仅针对多分量图像（如彩色图像）适用。JPEG2000

标准中定义了两种分量变换方式:可逆分量变换(Reversible Component Transformation, RCT)和不可逆分量变换(Irreversible Component Transformation, ICT)。RCT 和 ICT 都是去相关变换,通过对图像的三种分量进行相应的变换将图像变换到合适的彩色空间,提高整体编码效率。RCT 既适用于无损压缩,又适用于有损压缩;ICT 仅适用于有损压缩。

③ 小波变换。与 JPEG 相比, JPEG2000 最大的改进是以离散小波变换(DWT)代替了离散余弦变换(DCT)。离散余弦变换作为准最优变换,在图像处理中占据很重要的地位,但它有个重要的缺点是不具有时频局域性,考察的是整个时域过程的频域特征,或整个频域过程的时域特征。因此对于平稳过程,有很好的效果,但对于非平稳过程,就存在明显不足。在图像压缩领域表现更为明显,对于细节丰富、频率变化大的图像压缩效果差,另外,也无法实现 ROI 编码。而小波变换的最大特点就是具有良好的时频局域性,既能考察局部时域过程的频域特征,也能考察局部频域过程的时域特征,并且可以在高频时考察窄的时域窗,而在低频时考察宽的时域窗,因此不论是对于平稳过程还是非平稳过程,它都是强有力的工具,比 DCT 变换更适用于处理数字图像这样的非平稳信号。JPEG2000 标准中采用了两种小波变换滤波器组:有损变换(Daubechies 9/7 滤波器)和无损变换(Le Gall 5/3 滤波器)。这两种小波变换均可采用计算复杂度低的小波提升算法。

④ 量化。小波变换后虽然变换系数的个数并无减少(与原图像采样点个数相比),但信息分布发生了很大的变化,大部分能量集中在少数的小波系数中,利用量化可大量减少幅度很小的系数所携带的能量,提高整体压缩效率。量化主要是针对有损压缩进行,量化的关键是根据变换后图像的特征及重构图像质量要求等因素选取合理的量化步长。关于有损压缩量化步长的选取, JPEG2000 标准中有很大的灵活,没有给出统一的方法,用户可根据实际应用自行选择设计,而并不影响解码过程的进行,但需要把选取的量化步长作为参数放入比特流中传给解码器。量化步长与小波分解子带是一一对应的,一个子带一个量化步长,这样可充分利用人眼的视觉特性来提高编码效率。

⑤ ROI 系数处理。该模块也是可选模块,用于对感兴趣区编码。JPEG2000 中对 ROI 编码所采取的策略是:通过系数缩放使感兴趣区的数据位于更高的编码位平面(相对于背景区)。这样在编码比特流中, ROI 中的数据就排在背景区数据的前面。当压缩比较高时,可以优先保证 ROI 的恢复质量。

⑥ 熵编码。经分量变换、小波变换及量化后的图像数据,在一定程度上减少了空域或频域上的冗余度,但这些数据在统计意义上仍存在一定的相关性,为此采用熵编码来进一步消除数据间的统计相关。JPEG2000 熵编码过程可分为两个步骤:嵌入式码块编码和分层组织压缩位流。前者的基本思想是使压缩后的码流可划分为若干逐级包含的子集,每一子集表示对源图像的单元压缩;后者则可以使压缩码流具有质量上的可分级性,实现了图像的渐进式传输。

⑦ 打包-形成一定格式的压缩码流。为了适合图像交换, JPEG2000 标准规定了详细的语法结构来存放压缩码流及解码所需参数,以包为单位,形成最终码流。

### 6.1.3 压缩域图像检索原理

基于压缩域图像检索技术研究的关键问题是如何通过图像处理直接在压缩中提取图像的内容特征,即通过挖掘图像压缩时的中间结果或最终码流中包含的信息,力争在不解码或部分解码的情况下提取表征图像内容的特征,并以此作为索引实现基于内容的图像检索。

这是一个全新的研究思路，它将图像的压缩技术与检索技术融合在一起，避免了“全解压”的额外操作，大大提高了检索系统的实时性、灵活性和有效性。基于压缩域的图像检索系统的基本组成如图 6.6 所示，首先，它不需要特征库的技术，图像的特征可直接在压缩域中快速提取；其次，有利于动态图像数据的管理，不需要预先对图像数据进行处理（即提取特征，形成图像特征库）。因此，特别适用于互联网、Web 数据库的检索。

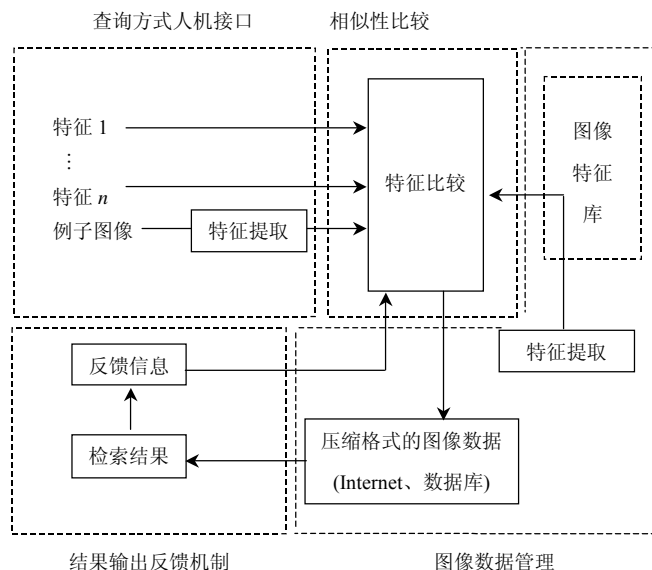


图 6.6 基于压缩域的图像检索系统的基本组成（黄祥林 2001）

面对图像压缩数据，传统的方法是将压缩图像进行全解码后在原始像素域进行处理，而在压缩域进行图像处理，就是对图像压缩码流不经解码或少量解码的情况下直接从图像压缩码流中提取图像的特征，即如何进行压缩域的图像处理，以得到图像的特征。图 6.7 给出了两种模式的比较示意图，可见基于压缩域的图像检索技术省去了一些中间环节。在压缩域的图像检索中，其提取数据的处理位置按压缩码流的可操作程度可分为三个位置。以变换编码的编解码过程为例，压缩域可操作的三个位置（黄祥林 2001）如下。

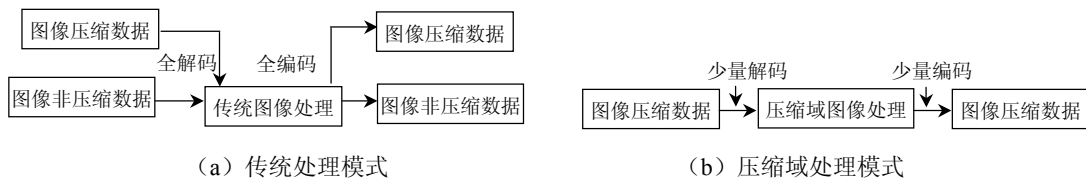


图 6.7 图像压缩数据的两种处理模式（黄祥林 2001）

(1) 熵解码前：这是压缩域处理的理想位置，但是经熵编码后所形成的压缩码流不具备结构化信息，其数据项并非字节对齐，不利于计算机的处理，所以，实际上在此处能进行的针对图像的操作较少，仅限于一些特殊情况，如利用修改 JPEG 码流的量化表来达到图像锐化的目的。

(2) 熵解码后：由于直接在熵解码之前进行图像处理非常困难，因此，压缩域的处理



操作一般要首先进行熵解码,再进行处理。在图像的编码过程中,熵编码是最后一道操作,其压缩能力为2到3倍,可见,对于具有25倍压缩比的JPEG图像来说,熵解码后还有10倍左右的压缩比率,其数据量与非压缩格式的数据量相比,仍然是较小的。

(3) 反量化后:有损编码时,一般都有量化步骤,因此在进行图像的高精度处理时,除了要进行熵解码外,还需要进行反量化,以提高图像处理的效果。在反量化后进行压缩域的图像处理也是常用的方法。

图6.8是一个典型的编解码过程图,位置0为像素域特征提取操作位置,可以看出,对于压缩图像,若在位置0提取特征,首先需要进行全解码。在压缩域操作中,所谓“不解码”,就是在位置1上未进行熵解码之前进行压缩域的处理,而“少量解码”,就是在位置2、3上进行熵解码或反量化之后进行处理。即使在位置2或位置3进行,计算量也可大大降低,反变换所需时间一般占总解压时间的40%~60%。

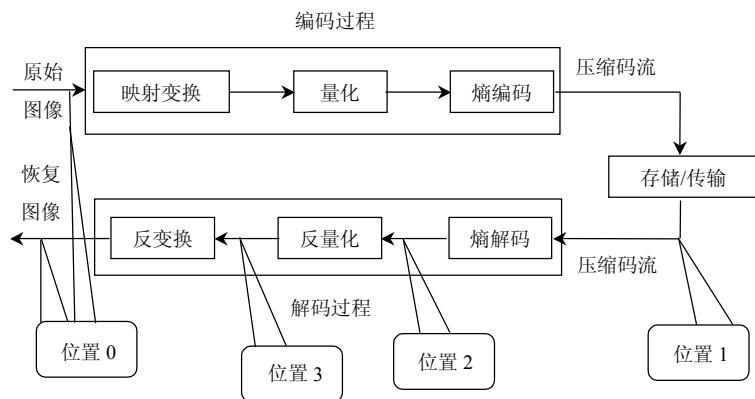


图 6.8 特征提取操作位置示意图（李晓华 2004）

#### 6.1.4 压缩域图像检索的研究内容

对已有压缩算法所形成的压缩数据的检索技术已经进行了许多年,也取得了一些成果,但由于现在的压缩算法在最初的设计实现时并未考虑后续图像检索处理的要求,所以在这些压缩数据上所能实现的图像检索的能力是十分有限的。一个根本的解决方法是从图像压缩的角度进行压缩域图像检索技术的研究,即进行支持检索的图像压缩方法的研究。本质上讲,图像压缩编码算法的主要目的是寻求图像数据的有效表征形式,这种表征不仅要具有小的存储要求,同时还能较好地保持原图像数据的信息含量。而从支持图像检索的角度来看,由于不同压缩算法所形成的压缩数据中包含了表征原始图像主要内容的大部分信息,所以理论上任何压缩码流都能对压缩域图像检索提供支持,而其区别主要在于压缩域图像检索处理过程实现的复杂度不同,且不同压缩算法对不同的图像检索操作具有不同的支持程度。因此,支持图像检索的压缩算法的研究就是要选择和实现有效的图像信号表征形式和量化方法,使其不仅有利于高效压缩的实现,同时还能够从其所形成的压缩数据(不解码或尽量少解码)中得到或构建出图像内容的特征量用于支持图像检索操作。总的来说基于压缩域的图像检索技术研究内容主要有两个方面(黄祥林 2001):

(1) 从处理角度看,由于现有的编码算法没有考虑到压缩域的分析处理,因此必须深

入分析压缩域的特有性质,研究其压缩域究竟有什么处理能力,力争在不解码(或不完全解码)的情况下进行处理,以提取特征信息进行图像检索,由于目前的大量压缩数据的存在,这方面的研究具有非常现实的意义。传统的提取特征信息的处理技术比较成熟,一个很自然的想法就是如何在压缩域推导出与这些处理算法相对应的对等操作(向辉等 1999)。例如,对于传统的空域处理中的一些技术就很容易推导出在以 DCT 为基础的压缩域中的对等操作,只是把它变成频率域处理而已。但并不是每一种传统方法都能找到对应的压缩域模式,也不是压缩域的处理模式全是传统模式对应的翻版。人们还要根据压缩域的特点,研究更符合实际应用的压缩域特有的图像处理新算法。

(2) 从编码角度看,研究新一代的编码算法,使其既具有较高的压缩效率,又具有支持压缩域检索与处理的能力,即同时考虑编码与处理,在更高的层次上来解决图像/视频信息的管理和操纵(魏海 2000),这将是新一代编码的研究方向,具有前瞻性。

### 6.1.5 压缩域图像检索的研究方法

基于压缩域的图像检索技术是近十年来多媒体领域的一个非常活跃的新型研究方向,但是需要在压缩域直接进行分析处理,就受到压缩算法本身局限的影响,这就给研究工作带来了相当大的难度。针对压缩域的图像检索技术的研究,其方法可归结为两种:第一,寻求传统处理方法在压缩域的对等操作;第二,寻求压缩域的特有操作。

#### (1) 寻求对等操作

这里所说的对等操作,指的是能够通过严格的数学推导,将一个域的操作转换到另一个域的操作。由于传统的图像处理方法的处理对象是基于原始像素域的,因此,许多空域处理操作很容易在频率压缩域中找到对应的处理方法,如基本的标量加、标量乘、向量加、向量乘运算等。许多压缩域的对等操作,都可通过严格的推导来证明,但在压缩域寻求空域的对等操作的过程中,一定要看其算法的简易性、有效性和对计算资源的需求,其综合指标(主要指处理时间、处理效果、耗费资源等情况)至少要好于或近似等于“解码”+“传统处理技术”+“再编码”的传统处理模式,否则没有实际意义,这是研究基于图像压缩域检索技术的根本出发点。

#### (2) 寻求特有操作

并不是所有的处理算法在压缩域中都具有对等操作,因此,如何根据压缩域所具有的特点,寻求压缩域所特有的图像处理算法是另一研究途径。对已有的压缩算法,深入分析压缩域所具有的特点,设计出方便、实用、高效的处理方法;在设计新的支持压缩域处理的图像编码算法时,充分考虑到压缩域的特性,提供具有更多操作功能的压缩码流。寻求压缩域的特有操作是压缩域检索技术中研究的难点,也是重点。

基于压缩域的图像检索方法与原始图像所采用的压缩方法有密切关系,用什么方法压缩对其后的检索有重要影响。考虑到常用的压缩技术的特点和压缩域检索的要求,可将压缩域检索技术分为空间压缩域技术、变换压缩域技术和混合压缩域技术三大类,如图 6.9 所示(李晓华等 2003)。空间域方法主要包括矢量量化和分形方法;变换域方法主要是利用 DFT(离散傅里叶变换)、DCT(离散余弦变换)、小波变换和 KLT(Karhunen-Loeve 变换)来进行压缩域特征的提取。目前大多数的研究成果都集中在变换域上,尤其是基于小波压缩域和 DCT 压缩域。下面,我们详细地介绍常用的几种基于压缩域的图像检索方法。

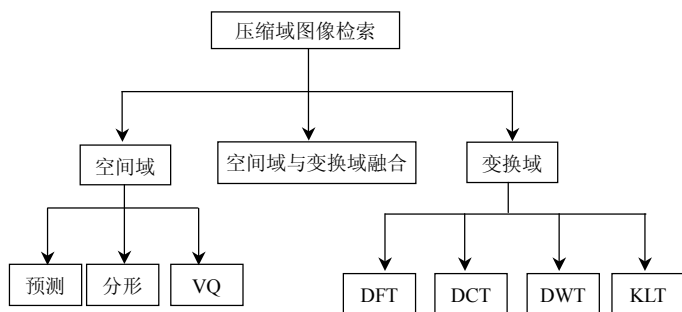


图 6.9 压缩域图像检索技术分类

## 6.2 空间压缩域

### 6.2.1 基于矢量量化

矢量量化是图像编码领域一个重要的编码方法。其中心思想是：通过训练集合首先构造一个有限集看做是码书，然后对任意输入图像中的每一分块矢量按一定的距离准则进行分类，并且将原图像中的所有分块矢量都采用与之相匹配的特征矢量码字来进行表征，而与所有分块矢量相匹配码字的索引序列就是最终的压缩数据。这样的编码结构和压缩数据非常适合与图像检索技术相结合，从结构上说，矢量量化本质上是一个聚类 and 分类的过程，这样的结构自然地与图像聚类分析以及分类处理等操作的要求相一致；从获得的压缩数据看，由于矢量量化是通过一个有限数量的再生矢量集合来对所有输入图像进行表征的，它的码字在一定程度上与图像的内容相对应，该向量集合也描述了解压缩图像的性质，所以对该矢量集合（码书）的任何分析处理操作都可以间接地反映到解压缩图像中。正是因为矢量量化具有的这种良好特性非常有利于图像的检索和查询，使得其可以广泛应用于图像增强、图像分类、边缘检测以及图像检索等方面（黄翔宇 2003）。

矢量量化的理论基础是香农的速率—失真理论。基本的矢量量化器可以定义为从  $d$  维欧氏空间  $R^d$  到其一个有限子集的一个映射，即  $Q: R^d \rightarrow C$ ，其中  $C = \{y_1, y_2, \dots, y_N | y_i \in R^d\}$  称为码书， $N$  为码书大小。该映射满足： $Q(x | x \in R^d) = y$ ，其中  $x = (x_1, x_2, \dots, x_d)$  为  $R^d$  中的  $d$  矢量， $y_s = (y_{s1}, y_{s2}, \dots, y_{sd})$  为码书  $C$  中的码字并满足

$$\text{dist}(x, y_s) = \min_{1 \leq j \leq N} \text{dist}(x, y_j) \quad (6-5)$$

$\text{dist}(x, y_s)$  是输入矢量  $x$  与码字  $y_j$  之间的失真测度（距离）。每一个矢量  $x$  都能在码书  $C = \{y_1, y_2, \dots, y_N\}$  中找到其最近码字  $y_j = Q(x | x \in R^d)$ 。输入矢量空间通过量化器  $Q$  量化后，可以用划分  $S = \{S_1, S_2, \dots, S_N\}$  来描述，其中  $S_i$  是所有映射成码字  $y_i$  的输入矢量集合，即  $S_i = \{x | Q(x) = y_i\}$ 。这  $N$  个子空间  $S_1, S_2, \dots, S_N$  满足

$$\bigcup_{i=1}^N S_i = S \quad \text{且} \quad S_i \cap S_j = \Phi (i \neq j) \quad (6-6)$$

基本的矢量量化编码和解码过程如图 6.10 所示。矢量量化器根据一定的失真测度（距

离) 在码书中搜索出与输入矢量之间失真最小的码字。在二次存储设备中存储时仅存储该码字的索引。矢量量化解码过程很简单, 只要根据码字索引在码书中查找该码字, 并将其作为重构矢量。内存容量能够容纳下码书时, 码书一般存储在内存中。矢量量化之所以能够压缩数据, 是由于它能够去除数据之间的冗余度, 有效的利用矢量中各分量之间相互有联系的四中性质: 线性依赖、非线性依赖、概率密度函数的形状和矢量维数的本身。

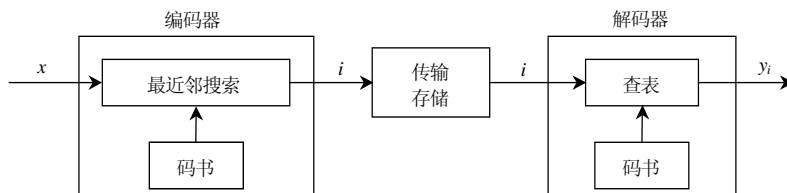


图 6.10 矢量量化中的编码和解码过程

从信源的角度来看, 矢量量化的结果可以认为是一个  $N-1$  阶有记忆信源所发出的一个符号集合, 每个符号是一个具有统计关联特性的  $N$  维矢量。对于相似的图像所对应的符号集合来说, 同一符号的概率分布应接近。在矢量量化过程中, 任一输入图像的分块矢量都依据统一的距离准则映射到有限特征矢量集合 (公共码字) 的某一码字上, 所以不同的图像与该有限矢量集合之间映射的统计特性在一定程度上反映了图像的内容。Idris (Idris F. et al 1995a) 基于这样的思想, 使用频数统计直方图构造了图像特征矢量集合中的矢量, 并将其作为表征图像内容的特征量, 用于比较查询图像和图像库中目标图像的相似度, 从而实现了将矢量量化压缩数据直接用于支持图像检索的目的。同时, 为了进一步降低对直方图特征量的存储要求以及直方图相互间距离计算的复杂度, 他们还对码字使用频数直方图进行了简化处理, 提出使用图特征量来对码字进行简化处理的方法 (Idris F. et al 1995b), 即某一码字矢量, 如果在图像压缩过程中使用过, 则与该码字对应的特征位置为 1, 否则为 0, 而不同图像间相似程度的计算只需对其使用不同的图特征量按位进行异或操作, 即可大大减小特征向量的存储量和距离计算的复杂度, 但这样简化处理的代价是降低了图像检索的正确率。图 6.11 和图 6.12 分别给出了上述两种统计特征量的生成示意图。

参考文献 (Barbas J.S. et al 1993) 探讨了大型雷达图像数据库的有效表征问题, 目的是为了优化存储量和搜索时间, 利用树状结构矢量量化的聚类方式, 并用树状结构矢量量化下的多分辨率小波来表征图像特征, 这种树状结构是根据脉冲的多分辨率小波分解结果

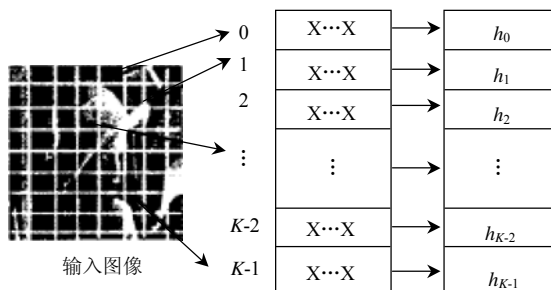


图 6.11 基于矢量量化的码字使用频数直方图特征构造示意图 (魏海 2000)

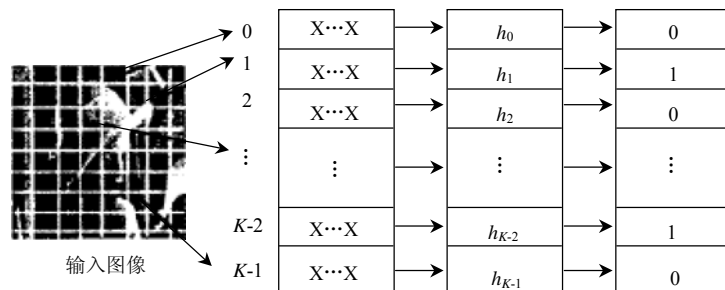


图 6.12 基于矢量量化的码字使用图特征量的构造示意图 (魏海 2000)

引入的, 因此该方法的整体性能较好。参考文献 (Vellaikal A. 1995) 将矢量量化技术应用到基于内容的遥感图像检索中, 采用估测多种形变的方式来提高作为内容描述符 VQ 码字的性能, 通过测试两种查询 (按类查询、按数值查询) 的性能, 发现后者的性能更为卓越。魏海等 (魏海等 2001a) 将矢量集中进行分类, 对不同类别的矢量进行不同的处理, 既具有较高的压缩比, 又能进行压缩域的图像检索。参考文献 (Eftekhari-Moghadam A.M. et al 2003) 设计了一种利用矢量表索引值提取图像连续区域直方图进行图像检索的方法, 具有较好的旋转、平移和尺寸不变性。

若根据图像中矢量的特性, 首先进行矢量分类。比如根据矢量方差值, 把矢量分成平滑矢量 (方差较小) 和细节矢量 (方差较大), 不同类矢量对应不同的码本。矢量量化时根据矢量的类型到相应的码本中寻找其码字索引, 这种矢量量化称为分类矢量量化。作为一种特殊的矢量量化方式, 分类矢量量化是由 Ramamurthi B. 所提出的 (Ramamurthi B. et al 1986), 其主要目的是为了克服在高压比下矢量量化所出现的边缘退化失真现象。其主要思想是: 首先对图像的分块矢量集合按一定的特征进行分类, 不同类别的矢量块子集根据其对于重构图像质量的重要程度分别采用不同长度的类别码书进行矢量量化。这样可以保证那些对人眼观察较重要的类别矢量块能够获得足够高的量化精度, 从而在总体上能够获得主观质量更好的解码图像。基于分类矢量量化的检索方法与上述基于矢量量化的方法类似, 不同之处仅在于分别统计各类码本中码字矢量使用频数的统计直方图, 然后把它们连成一个统一的直方图。基于分类矢量量化进行检索有一个优点, 可以根据人眼对各类矢量的敏感程度, 对直方图中的分量采用不同的权值来提高检索效率。魏海等比较了标量量化 (SQ)、矢量量化 (VQ) 和分类矢量量化 (CVQ) 用于检索的性能 (魏海等 2001c)。结果表明, VQ 的检索性能优于 SQ, 而 CVQ 的检索性能明显优于 VQ。

### 6.2.2 分形编码

分形的理论基础是法国数学家 Mandelbrot 所创立的分形几何学 (Mandelbrot B. 1982)。具体关于分形的定义及理论在第 5 章中已做过详细介绍, 这里重点论述分形编码的思想。

对于二维的灰度图像, 其分形压缩变换可定义为

$$W_i(R_i) = s(R_i)S^{(n+1)}\tau(R_i)D(R_i) + o(R_i) \quad (6-7)$$

式中,  $D(R_i) \in \{D_1, D_2, \dots, D_m\}$  是域块;  $\tau(R_i) \in \{\tau_1, \tau_2, \dots, \tau_8\}$  是 8 种对称变换;  $S^{(n+1)}: \mathfrak{R}^{2^{2(n+1)}} \rightarrow \mathfrak{R}^{2^{2n}}$  是下抽样算子;  $s(R_i) \in \{s_1, s_2, \dots, s_n\} \subset \mathfrak{R}$  是缩放系数;  $o(R_i) \in \{o_1, o_2, \dots, o_l\} \subset \mathfrak{R}$  是偏移量,  $R_i$  是值块。

由以上定义,可以看出分形是一类无规则、混乱且复杂,但其局部与整体具有相似性的系统,所以分形最主要的特征就是其所具有的自相似性。这种自相似性既可以是形态或结构几何意义上的相似,也可以是信息或功能统一意义下的相似。而在实际中,具有统计意义下的自相似分形集合占大多数(魏海等 1998)。除了具有自相似性这一特性外,分形集合所具有的另一个诱人之处在于对大多数的分形集合来说,尽管表面很复杂,但都能够从少数参数出发,按极简单的规则来生成它,也正是这一点促使人们将分形应用于图像的压缩编码中。所有的图像都可以认为是具有自相似性的分形系统,如果我们能够对一幅图像找到它所对应的生成规则和参数,这些参数和规则便唯一地表示了这幅图像,而这些参数和规则所具有的数据量相对于原始图像的数据量而言是极其微小的,这样就能够获得极高的压缩比。当然,严格来说,我们不可能绝对得到任一幅图像所对应的分形规则,但至少在统计的意义下我们可以选择一些规则来得到一幅图像的分形近似图像来足够好地接近于原始图像。正是由于这些原因,分形理论一经引入图像压缩领域便引起了众多研究人员的普遍关注,相继出现了许多令人注目的分形编码方法。

另外还有一种迭代分形编码,其理论基础是分形几何学中的迭代函数系统理论,迭代函数系统 IFS 通常是指:在某一度量空间内的收缩仿射变换集  $W\{w_i|i=1,2,\dots,N\}$ ,它是一种通过寻找信号自身递归的变换不变关系来构造自相似分形的方法。根据 Banach 收缩映射不动点定理,只要存在常数  $0 < s < 1$ ,使得

$$d(w_i(x), w_i(y)) < sd(x, y), \forall i, x, y \quad (6-8)$$

就存在唯一的“吸引子集”  $A$  满足“自相似性”,即

$$A = W(A) \underline{\Delta} \bigcup_i w_i(A) \quad (6-9)$$

和“吸引性”:任何集合  $B$  经过收缩仿射变换算子集  $W$  的反复作用都会收敛到  $A$ 。同时,不论从  $B$  中的哪一点开始,只要按同样的顺序嵌套 IFS 中的变换,都会收敛到  $A$  中的同一点。

分形编码算法主要利用了图像内部的自相似性,具有很高的压缩效率,但由于编解码巨大的不对称性,使其应用受到了限制。在基于块的分形编码算法中,输入图像被分割成互不重叠的 Range 块,对每个 Range 块,选择一个 Domain 块,这个 Domain 块能通过一些仿射变换来最大限度地相似对应的 Range 块,这些仿射变换就是分形码,它以高压压缩的形式表征了这幅图像。可见,分形编码实际上是一种自我索引技术,基于分形编码的索引算法的研究主要就是利用了这一特点。

Sloan (Sloan A.D. 1994) 首先发现了分形图像编码技术在基于内容的图像检索中的应用前景,并进行了探索,提出了将图标图像数据库中的图像直接拼凑起来进行编码的方法,即对于任何图像中的排列块可以在任何图像中寻找区域块;而两幅图像之间的相似性则可以用一个图像中的子块当作另一个图像中区域块的数目来度量。对这方面的进一步研究是由 Zhang 等人开展的。Zhang 等人 (Zhang A. et al 1995) 提出直接利用图像的分形码来实现基于纹理的图像检索技术,这里分形码是以图像子块为单位得到的,即首先把图像分成不重叠的若干子块,对每一子块根据分形原理寻找其分形码,检索时直接比较查询图像和目标图像的分形码。参考文献 (Zhang A. et al 1996) 提出了一种基于九分树分解和分形编码的图像检索策略,并进一步比较了基于分形的方法与小波子带能量法的检索效果。结

果表明,对于纹理图像,小波方法要好些,而对于其他图像,则分形方法能略胜一筹。Ida (Ida T. et al 1995) 还提出了一种基于分形码的图像分割方法,主要是基于以下设想:如果  $S$  区域中有 Domain 块,则其相应的 Range 块也在这个区域内,因此可以由分形码分割出区域  $S$ ,这完全可用到图像检索技术中。参考文献(魏海等 2002) 利用分形对图像所具有的独特表征能力,对基于迭代分形的压缩方法和检索方法进行了研究,在小波变换域内提出了一种基于迭代分形的高效图像压缩方法,并在图像分形码的基础上,利用迭代函数系统的分布特性来进行图像间相似距离的计算,从而实现了基于迭代分形压缩数据的图像检索操作。参考文献(杨守义等 2003) 利用图像的分形编码数据里包含的空间域信息,提出了一种在分形域基于内容的图像检索,该方法在分形编码的基础上定义了一个提取图像与偏移量图像,通过计算两者的距离来求原始图像和目标图像的相似度。魏海等(魏海 2000) 对基于迭代分形的图像压缩和检索方法进行了初步的研究,提出了一种基于迭代分形的高效图像压缩方法,并在图像分形码的基础上,利用迭代函数系统的分布特性来进行图像间的相似距离的计算,从而实现了基于迭代分形压缩数据的图像检索操作。

### 6.2.3 预测编码

预测编码的基本思想是:依次扫描图像中各个像素,通过提取每个像素中的新信息然后对这些信息编码来消除像素中的冗余。这里像素的新信息定义为该像素的当前值与预测值的差。

预测编码可分为无损预测编码和有损预测编码。无损预测编码包括两个步骤:预测和编码。预测时,逐个读取输入图像像素序列  $f_n(n=1,2,\dots)$  中的每个像素,预测器根据若干个过去的输入产生当前输入像素的估计值。例如,可通过将  $m$  个先前的像素进行线性组合以得到预测序列,即

$$\hat{f}_n = \text{round} \left[ \sum_{i=1}^m a_i f_{n-i} \right] \quad n=1,2,\dots \quad (6-10)$$

式中,  $m$  是线性预测器的阶;  $\text{round}$  是舍入函数;  $a_i$  是预测系数。由于图像内像素的相关性,预测序列与实际序列的差,即预测误差序列  $e_n(n=1,2,\dots)$  的概率密度函数一般在零点有一个高峰。对预测误差序列再用符号统计编码的方法编码可以得到压缩的结果。

一般对预测编码图像的检索方法有两种,一种仅将编码结果解码(部分解码但不变动预测结果),另一种直接利用编码结果(完全不解码)。

① 仅将编码结果解码的方法。如果仅将编码结果解码,得到的是预测误差序列。对每个像素来说,对它的预测误差是对它邻域的一种统计描述,即该像素相对其邻域的变化量。因为图像中邻域的性质与纹理有密切关系,则可以认为预测误差在一定程度上反映了纹理信息,可以用来进行基于内容的图像检索。

② 直接利用编码结果的方法。在无损预测编码的第二个步骤中常用统计(符号)编码方法,如哈夫曼方法等,统计编码的特点是把较短的码赋予出现概率较大的预测误差值,而把较长的码赋给概率较小的预测误差值。具体来说,一个哈夫曼码本身就包括每个预测误差值所赋的码字,而每个码字的长度与它所赋予的预测误差值的出现概率成反比。所以哈夫曼码本中也基本包含了预测误差值直方图中的信息,可以用来进行基于内容的图像检索。

## 6.3 变换压缩域

许多图像编码方法的基本技术都是以某种数学变换为基础的,如 DFT 变换, K-L 变换, DCT 变换及小波变换等,不少学者对这些变换压缩域中的图像检索技术进行了深入的研究。

### 6.3.1 基于DFT变换域

DFT 变换是图像处理和信号分析处理中的一个非常重要的数学变换,由于 DFT 利用复数指数函数进行变换,其能量压缩特性良好,从而可提供良好的编码性能。DFT 的某些性质在检索和模式匹配中是很有用的,和图像检索、匹配有关的重要性质有:

- (1) 系数幅值的平移不变性;
- (2) 简化空域的相关运算等。

DFT 正反变换的定义如下:

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk} \\ x(n) &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}nk} \end{aligned} \quad (6-11)$$

Stone 等人提出了一种在 DFT 压缩域的图像检索算法 (Stone H.S. et al 1996), 这种算法合理组织傅里叶系数, 同时, 该算法提供两个阈值, 用户可通过这两个阈值来独立地调整匹配的近似度, 使用一个阈值来控制亮度匹配, 再用一个阈值控制纹理的相似程度, 用这两个阈值可以使匹配达到由粗到精的过程。另外, 由于这两个阈值之间是相关的, 故利用傅里叶系数可以有效地计算出图像的特征, 且在傅里叶系数大部分是零时, 这种计算尤其有效。参考文献 (Augustejin M. et al 1995) 等利用 DFT 系数的统计特性和 DFT 系数径向分布结合角度分布实现了卫星图像的纹理分类。其中的统计特征有最大幅度、平均幅度、幅度的能量和傅里叶系数幅度的方差等。同时, 还研究了基于傅里叶系数半径 (即幅度) 和相位分布特性的检索性能。DFT 系数的径向分布在一定程度上反映了图像纹理的粗糙程度, 而 DFT 系数的角度则反映了图像纹理的方向性。另外, 文中还对比了两种方法的纹理分类性能, 结果表明当图像中存在明显突出的频率成分时, 基于 DFT 系数径向和角度分布的方法具有较好的分类效果, 而当频率成分比较均匀时, 基于统计的方法则具有更好的分类效果。

使用极坐标下的傅里叶系数分布来进行纹理图像的分类也是常用的算法之一。参考文献 (Celentano A. et al 1997) 研究了基于 DFT 系数角度分布的图像检索方法。先对图像进行低通滤波的预处理; 然后进行傅里叶变换; 在  $180^\circ$  范围内扫描 DFT 频谱图, 统计每一角度范围内的中频 DFT 系数之和, 得到一个角度分布直方图, 以此作为特征矢量进行图像检索。虽然该特征矢量具有平移不变性, 但是在图像发生旋转后, 它将发生循环移动, 因此该方法不具有旋转不变性, 对这一缺点可通过对特征矢量进一步作 FFT 变换使其具有旋转不变性。

虽然基于 DFT 压缩域的检索技术取得了较好的结果, 但是由于 DFT 是一个复数变换, 运算量大, 实现困难, 另外对于一般的图像信号, 利用 DFT 进行压缩, 效果也不是很理



想, 因此很少被用来进行图像压缩, 这也是图像压缩标准中没有采用 DFT 变换的原因。因此, 基于 DFT 压缩域的图像检索能力也就无法很好地体现。目前仍有学者在做这一方面的研究, 但是只是利用 DFT 作为特征提取的一种手段, 不再强调压缩域检索这一概念(张问银 2005)。

### 6.3.2 基于DCT压缩域

DCT 变换是从 DFT 演变而来的, 对许多自然图像来说, DCT 变换接近最优的 K-L 变换, 是现在应用最广泛的多媒体数字压缩技术, 图像、视频压缩国际标准 JPEG、MPEG 和 H.261 都采用了 DCT 技术。二维离散余弦变换的定义如下:

$$F(j,k) = a(j)a(k) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m,n) \cos \left[ \frac{(2m+1)j\pi}{2N} \right] \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \quad (6-12)$$

$$f(m,n) = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} a(j)a(k) F(j,k) \cos \left[ \frac{(2m+1)j\pi}{2N} \right] \cos \left[ \frac{(2n+1)k\pi}{2N} \right] \quad (6-13)$$

式中,  $a(k) = \begin{cases} \sqrt{1/N} & \text{如果 } k=0 \\ \sqrt{2/N} & \text{如果 } k=1,2,\dots,N-1 \end{cases}$ ,  $N=8$ ,  $F(j,k)(j=1,2,\dots,8, k=1,2,\dots,8)$  是

DCT 系数,  $f(m,n)(m=1,2,\dots,8)$  是输入二维数据的采样。

DCT 变换有许多优点。首先, 它是一种线性正交变换, 变换核各矢量间单位正交, 可以将  $8 \times 8$  图像的空间域转换为频率域, 只需要用少量的数据点表示图像; 其次, 二维 DCT 变换是对称分离的, 即二维 DCT 变换可分解为行列方向的两次一维 DCT 变换, DCT 有快速算法, 算法性能很好, 易于在软硬件中实现, 而且 DCT 算法对称的, 利用逆 DCT 算法可以用来解压缩图像; 最后, DCT 产生的系数很容易被量化, 因而能获得好的块压缩, 性能优于 DFT 等其他变换, 去相关压缩能力接近于 KLT 最佳变换。更为重要的原因是, 经过 DCT 变换后所得 DC 系数和 AC 系数分别具有以下特性:

- (1) 对于每个变换块, 每一个分量(直流 DC 分量)表达了此块的平均亮度信息;
- (2) 低频信息集中于变换块的左上角, 高频分量位于变换块 Zigzag 排序(Zigzag 排序描述了  $8 \times 8$  DCT 块的频率的增长走向)后的后方(DCT 变换实际上是空间域的低通滤波);
- (3) 交流(AC)系数反映了像素间差异的频率信息和方向信息。

这些都决定了利用 DCT 的特征提取技术进行图像检索具有很大的吸引力和可行性, 也是近年来国际上有关领域的研究热点之一。

基于 DCT 压缩域的检索技术就是在现有的压缩标准的基础上, 通过分析 DCT 变换系数以及压缩算法特点, 力争在不解码或部分解码的情况下实现图像检索。利用 DCT 良好的数学变换性质能够在 DCT 域进行高效的图像检索, 避免了传统空域操作烦琐的解 DCT 压缩数据, 处理后再还原为 DCT 压缩数据的操作过程, 将其应用于图像检索能够充分发挥检索准确率高, 检索速度快, 计算复杂度低, 计算量小等优势。

在 DCT 压缩域进行图像检索, 主要是分析、提取能反映图像内容的 DCT 系数的特性。目前大致有两类方法(沈兰荪等 2000): 第一类方法是对 DC 图进行处理, 因为 DC 图是整幅图像的缩略图, 表达了图像的平均能量, 所以可采用传统的图像检索方法(如灰度直方图等), 但是没有利用 DCT 域的低中频信息; 第二类方法根据交流(AC)系数反映了像

素间差异的频率信息和方向信息,通过计算 AC 系数的能量直方图等信息来检索图像,其检索效果并不太理想。应该说目前对基于 DCT 压缩域的图像检索技术研究还处于比较初级的阶段,还只是对 DCT 数据的简单直观应用,如 DCT 系数的三个大致方向性分布特征,初步判定边缘的存在与否和边缘方向;用直流系数的统计直方图作为图像相似性测度和判断等。这里既有现今基于内容的描述和检索正处于探索阶段的原因,也有 DCT 系数  $8 \times 8$  分块特性在某种程度上并不利于对整体信息表达的原因。

基于 DCT 压缩域的检索技术中研究最多的是纹理特征的提取及基于纹理特征的检索和分类。Smith (Smith J.R. et al 1994) 首先提出一种基于 DCT 的纹理图像检索方法,它将图像分成  $4 \times 4$  的子块,对其进行 DCT 变换,对所有块在同一位置上的系数计算其均值和方差,构成一个具有 32 个分量的特征矢量,用它来表征整个纹理的图像特征,并用 Fisher Discriminate Analysis (FDA) 来进行降维处理。这种方法由于采用的 DCT 块大小和压缩标准不兼容,要使用这种方法时,必须完全解码才能提取特征,但它给我们提供了一种 DCT 压缩域图像检索的思路。Reeves (Reeves R. et al 1997) 在此基础上进行改进,采用与标准兼容的  $8 \times 8$  子块,并且认为反映不同纹理特性的主要是前几个 AC 系数,从而避免了判别分析,它只对每个块内的前八个系数计算方差来形成特征矢量,由于这种方法生成的特征向量维数比上一种方法少,所以它的算法时间复杂度也比上一种方法小。与上述计算单个通道均值和方差的策略不同,参考文献 (Bae H.J. et al 1997) 根据 DCT 系数的特性,把 DCT 系数分成不同的区域,通过统计这些区域的能量得到一个具有 9 个分量的综合特征。Sim 等人 (Sim D.G. et al 2001) 引入了人眼视觉特性,利用一个符合人眼视觉掩模对各个频率通道的特征分量做加权处理。参考文献 (Furht B. et al 1998) 等考虑到直方图技术在图像检索中的作用,提出了一种基于低频 DCT 系数能量直方图的检索方法。该方法首先选择 DCT 块中的一个对称区域,然后统计该区域中的 DCT 系数的能量直方图,并以此作为索引进行图像检索。

上述各种方法所提取的纹理特征抗干扰能力通常不是很强,当图像由于压缩倍数较高或有噪声的情况下,检索及分类效果往往不好。为此,黄祥林 (黄祥林等 2002a) 适当组织 DCT 系数使之反映图像纹理的方向性,提出了一种基于 DCT 区域能量方向性的纹理图像分类方法。首先根据一定区域的变换系数代表着一定方向的频谱成分,将  $8 \times 8$  子块 DCT 变换系数分成直流区、竖直纹理区、水平纹理区和对角纹理区,如图 6.13 所示。四个区域的能量直接从 RLE 码流中采用式 (6-14) 来计算,即

$$\begin{aligned} E_0 &= \sum_{m=1}^N \left( A_{00}^m - \bar{A}_0 \right)^2 \\ E_k &= \sum_{m=1}^N \sum_{\Omega_k} \left( A_{ij}^m - \bar{A}_k \right)^2 \quad (k=1,2,3) \end{aligned} \quad (6-14)$$

式中,  $\bar{A}_0 = \frac{1}{N} \sum_{m=1}^N A_{00}^m$ ,  $\bar{A}_k = \frac{1}{21N} \sum_{m=1}^N \sum_{\Omega_k} A_{ij}^m$  ( $k=1,2,3$ )  $\Omega_k$  ( $k=0,1,2,3$ ) 分别表示原图像中直流成分区域、水平方向的区域、对角方向的区域和垂直方向的区域,并且  $\Omega_k$  中的系数个数为 21。

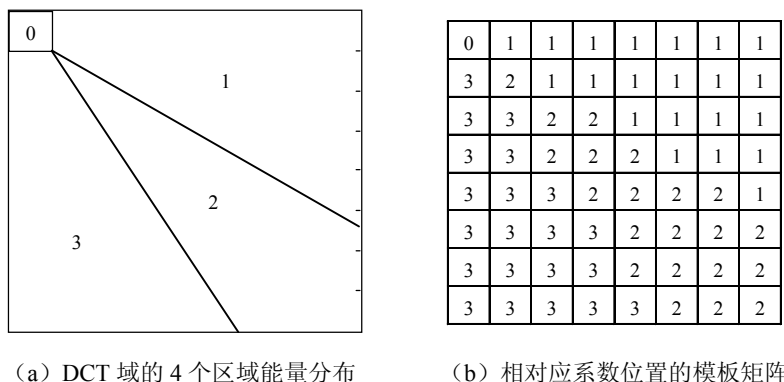


图 6.13 DCT 区域的能量分布 (黄祥林, 2001)

最后通过一种多分辨率特征组织形成特征矢量来支持纹理分类及检索。实验结果表明这种方法不仅计算简单, 而且在有噪声的情况下也具有很好的检索结果。黄祥林还提出一种基于 DCT 频率系数分布特点的纹理分类方法 (黄祥林等 2002c): 对  $8 \times 8$  子块 DCT 变换系数按低频率、中频率和高频率分区, 从 RLE 码流中直接统计这些区域的能量, 形成多分辨率特征矢量。由于这里是把各个方向中同等地位的系数考虑在同一个区域, 因此这种方法具有较好的放置不变性。这两种方法的另外一个特点是结合 JPEG 压缩算法的具体特点, 直接从 RLE 码流中计算特征矢量, 真正实现了基于压缩域的图像检索。Fan (Fan Y. et al 2002) 等直接在 DCT 域提取有方向性的图像特征进行检索, 并指出 DCT 域图像特征信息还需要进一步的开发和利用。Feng 等利用归一化的区域面积向量进行 JPEG 图像的检索 (Feng G.C. et al 2003)。Climer 等 (利用 DCT 系数中的 DC 系数和少量重要的 AC 系数形成图像的四叉树索引结构 (如图 6.14 所示), 然后采用逐层相似性比较的方法, 加快了检索速度 Climer S. et al 2002)。

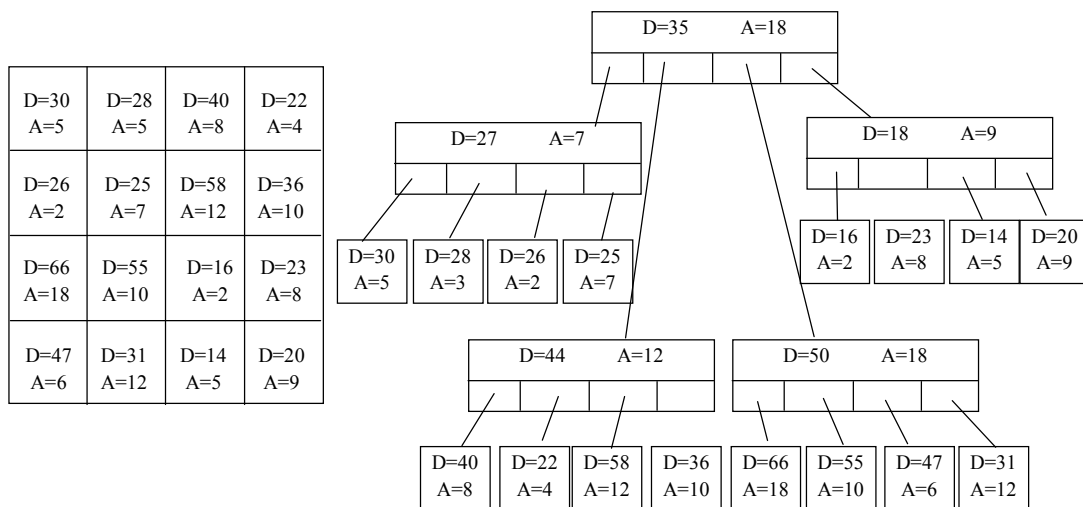


图 6.14 四叉树构造图 (张问银 2005)

图 6.14 中的四叉树构造方法只利用了块中的 DC 系数和块中最大的 AC 系数, 四个相

邻的块分别作为子结点，又构造出一个父结点，父结点中的  $D$  值是四个子结点的  $D$  值的取整均值， $A$  值则是子结点中  $A$  值的最大值。相邻的父结点又类似地构造出祖父结点，直到将所有的块全部组合成一棵四叉树，这种方法具体的距离计算方式如下：

$$D = \sum \left[ (D_q - D_n) + \left( \frac{A_q - A_n}{\lambda} \right) \right] \quad (6-15)$$

$D_q$  和  $D_n$  分别是被查询图像和查询图像对应结点的平均 DC 值， $A_q$  和  $A_n$  分别是相应的最大 AC 系数。相似性比较时，从根结点开始逐层比较，对应结点的累加值小于预设的门限值，就认为是相似的。当图像尺寸很大时，四叉树也很大，当将全部结点比较完时，很费时间，所以不一定要将所有的结点全部比较完，否则运算量很大，同时，四叉树的存储会占用很大的空间，这种方法对图像亮度的改变比较敏感。

Jose 等 (Lay J. A. et al 1999) 利用 DCT 系数的量化值去构造能量直方图，提出一种 DCT 系数量化直方图算法。量化后的 DCT 系数是一个整数，其最大值是直方图的维数长度。一个  $8 \times 8$  块的能量直方图  $H_c(m)$  表示如下：

$$H_c(m) = \sum_u \sum_v \begin{cases} 1, Q(F(u,v)) = m \\ 0, \text{其他} \end{cases} \quad (6-16)$$

式中， $Q(F(u,v))$  表示 DCT 系数在  $(u,v)$  的量化值，量化过程如下：

$$Q(F(u,v)) = \text{IntegerRound} \left( \frac{F(u,v)}{Q(u,v)} \right) \quad (6-17)$$

量化系数能量直方图体现了不同的能量强度在所有的块中所发生的次数。不同图像有着不同的能量强度水平，相似的图像有可能在能量强度层次上相似，Jose 的这种方法就利用了这一点。它的相似度衡量方法为归一化的 L1 距离方法，即

$$D(H_q, H_c) = \frac{1}{n} \sum_{i=1}^n \frac{|h_i^q - h_i^c|}{M} \quad (6-18)$$

式中， $H_q$  和  $H_c$  分别是被查询图像和查询图像的量化系数直方图， $n$  为直方图的维数。 $M$  是在某个能量强度上系数发生的最大次数。这种方法不需要反量化，可直接利用量化的 DCT 系数作统计，比四叉树方法要简洁得多，不过事先要确定直方图的长度，即最大能量强度。

除纹理特征外，边缘信息在表征图像内容上也具有重要意义。Abdelmalek (Abdelmalek A.A. et al 1994) 提出了利用 DCT 系数检测方向线的技术。利用空域坡度为  $m$  的直线在 DCT 域中表现为坡度大约为  $1/m$  的直线这一特性，在 DCT 域中提取水平、对角、垂直方向的特征信息进行图像的匹配。Shen (Shen B. et al 1996) 提出了一种从 JPEG 的 DCT 高频系数中检测出图像边缘和重要区域的方法，其中，边缘方向包括水平、垂直、对角、基本水平和基本垂直 5 种方向。实验结果表明，基于 DCT 的边缘检测有着和 Sobel 边缘检测算子相媲美的性能。Lee (Lee S.W. et al 2000) 提出了一种利用边缘信息进行图像匹配的方法，该方法首先从 AC 系数中提取二值边缘图，然后利用边缘图计算边缘方向、边缘强度和边缘偏移量形成特征矢量，最后按一定的相似性准则进行图像匹配。实验结果表明该方法具有较好的边缘检测能力和检索效率。黄祥林 (黄祥林等 2002b) 等参考三级小波分解的情况 (如图 6.15 所示)，将 DCT 系数进行重组，得到 DCT 域的多分辨率特性 (如图 6.16 所示)。

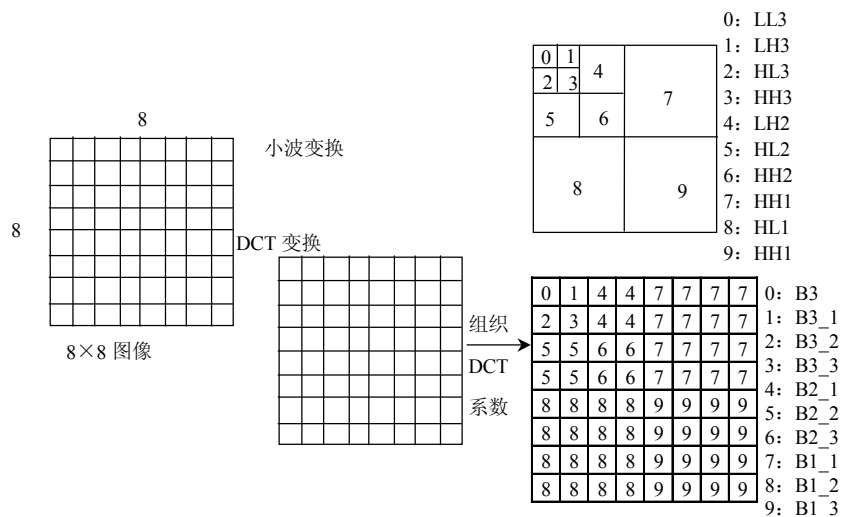


图 6.15 8×8 图像的小波变换与 DCT 变换 (黄祥林 2001)

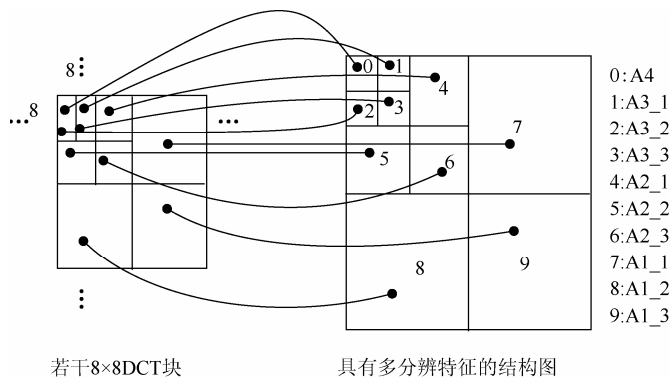


图 6.16 整幅图像中 DCT 块系数区域的组织 (黄祥林 2001)

在此基础上, 设  $(x, y)$  表示图像区域中像素位置的坐标, 将  $A2\_1, A2\_2, A3\_3$  区域内同一位置的系数中的最大值组成一个与  $A2\_1$  区域同样大小的  $I(x, y)$  图像块, 即

$$I(x, y) = \max \{A2\_1(x, y), A2\_2(x, y), A2\_3(x, y)\} \quad (6-19)$$

然后对  $I(x, y)$  图像块进行简单的量化处理, 即

$$\begin{aligned} M1 &= \max \{I(x, y)\}, \\ M2 &= \min \{I(x, y)\}, \\ \text{设 } M3 &= \text{Ave} \{I(x, y)\}, \\ M &= (M1 - M2) / 3 \end{aligned} \quad (6-20)$$

则有:

如果  $I(x, y) > M3 + M$  或者  $I(x, y) < M3 - M$  那么

$$I(x, y) = 255$$

否则  $I(x, y) = 0$

经过这样对  $A2$  区域的处理就得到了图像的大致轮廓  $I(x, y)$ , 对  $A3$  区域也进行同样的

轮廓提取,而对 A4 区域(即 DC 图)则采用简单的 Roberts 算子提取轮廓。当得到图像的大致轮廓后,进行基于轮廓的连通直方图的构造。对于一给定分辨率为  $X \times Y$  的轮廓图像  $I(x, y)$ ,先计算其连通区域,再计算具有相同连通区域面积(用像素点表示)的概率。连通直方图可通过式(6-21)计算;即

$$H(k) = \sum_{I(x,y)} \delta^k \quad (6-21)$$

这里,  $\delta^k = \begin{cases} 1 & \text{如果 } C = k \\ 0 & \text{否则} \end{cases}$

$k$  表示连通区域的量化面积,  $k=1,2,\dots,M$ ;  $M \leq (X \times Y)/S$ ,  $S$  是连通区域面积的量化步长,  $C$  表示某个连通区域的量化面积。在实际应用中,为消除图像尺寸的影响,采用归一化的连通直方图,即

$$H'(k) = \frac{H(k)}{\sum_k H(k)} \quad (6-22)$$

除了上述的纹理特征和边缘信息的提取,研究者还提出了其他的基于 DCT 域的图像检索方法。参考文献(张问银等 2004)定义某 AC 系数与该块中 DC 系数的比值为该系数的 A/D 能量,在此基础上给出了一种基于 A/D 能量直方图的 JPEG 图像检索算法,直接在 DCT 域中计算能量直方图,检索效果显著。参考文献(李鹏杰等 2004)分析了 DCT 块中系数分布的特点,提出了一种 DCT 域中 MPEG-7 主色描述符的提取算法,提高了压缩域图像进行特征提取的速度和效果。参考文献(张问银 2005)直接在压缩域中利用 DCT 系数进行块分类,每一类分块形成一个二值索引图,统计该索引图的归一化转动惯量(Normalized Moment Inertia, NMI)值作为该类的一个特征,所有类的 NMI 特征构成了图像的一个特征序列,以此进行图像检索,该方法不需要完全解压缩,降低了计算复杂度,对图像的平移、旋转和尺度变换有较好的鲁棒性。Shneier 等人提出了一种基于 JPEG 压缩域的图像检索算法(Shneier Michale, et al 1996),其主要思想是:通过判定检索图像和目标图像中不相连区域对中 DCT 系数相似关系的大小进行检索,先在图像内选取 2K 个互不相连的区域窗,随机配对得到 K 个窗对,对每个窗所包含的  $8 \times 8$  子块中每个系数进行均值计算,得到 64 维的特征矢量,将配对窗的特征矢量之间按其对应关系判定每一对分量的相似度大小并赋予该窗对一个比特(0 或 1),而检索图像和目标图像的相似与否取决于所有这些窗对比特流的相似度。Yu (Yu Hong Heather 1999)提出一种可用于直接比较两幅 JPEG 压缩图像相似度的测度, Q-测度,可直接利用 DCT 系数计算两幅图像之间的相似程度。

Chang (Chang C.C. et al 2004)等在 DCT 域中设计了一种提高检索效率的 JPEG 图像检索方法,称为 DCT 系数相关法。它根据 DC 和 AC 系数之间的相邻关系构造图像的索引特征,分别称为 DC 特征和 AC 特征,以此作为相似性检索的依据。具体特征的构造过程如图 6.17 所示。DC 特征是根据相邻块的 DC 系数的差值决定的一维 0、1 向量,差值大于等于 0,对应块被标记为 1,否则为 0,向量的长度等于块的总数。AC 特征是根据块中选取的 9 个 AC 系数相邻间的差值产生的,这些差值产生一个十进制索引值,由所有块的索引值产生一个直方图,它表示了一个索引值所对应的块数。两幅图像的 DC 特征的比较采用异或运算,直方图的比较采用最简单的 L1 距离方法。两种特征的相似性采用加权法确定最终的相似程度,小于这个程度的图像则认为所检索的结果。这种方法在特征

的构造上比较直接,运算量少,所以检索效率较高。但对于图像的旋转缩放变换,鲁棒性稍差。

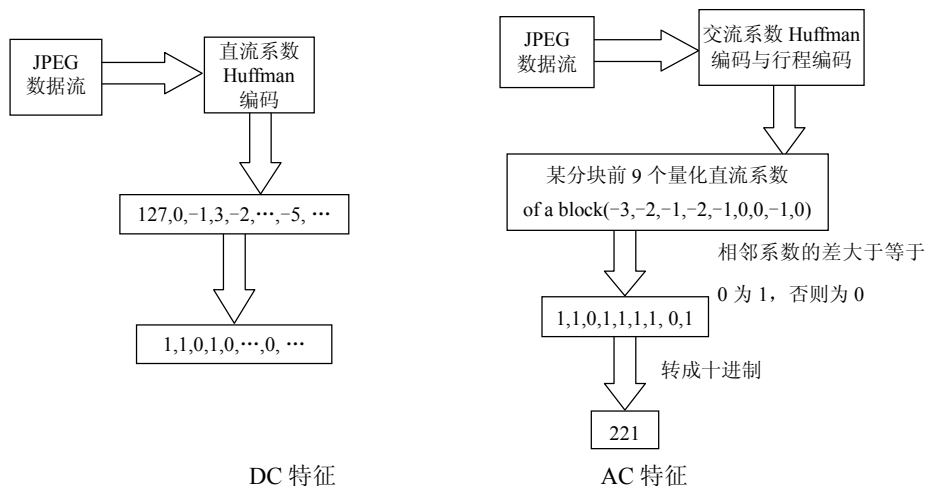


图 6.17 系数相关法的特征构造示意图 (张问银 2005)

作为语义特征,图像中的字符对于理解图像/视频的内容有很大的帮助。要获得这种语义特征,需要先定位图像/视频中的字符区域。Zhong 等 (Zhong Y. et al 2000) 提出一种 DCT 域字符区域定位方法,认为一行字符中字符与字符的间距对应于水平方向的频率变化,而多行字符之间的行距表现为垂直方向的频率变化,从而设计出一种两级字符区域定位方法:首先计算每一 DCT 子块的水平亮度变化信息,当亮度变化信息大于一定阈值时就认为该块为候选文本块,然后运用形态学法滤除孤立噪声块并合并不相连的文本块,最后再用垂直亮度变化信息确认文本区域。黄祥林 (黄祥林等 2002d) 等人也提出了一种 DCT 域的字符定位方法。这种方法依据字符具有特殊的线条结构而表现出明显的竖向、斜向、横向纹理特征的特点,计算每一 DCT 子块相应方向的频率变化程度,将其作为字符的纹理特征并结合阈值法检测字符区,试验表明,该方法的检测率可达 97% 以上。上述两种方法对尺寸较大且笔划为实心的字符区域及字符对比度较差的情况容易造成漏检。第一种情况是因为在  $8 \times 8$  的子块中显示不出其频率的变化特性;后一种情况则可通过设计自适应阈值来解决。

### 6.3.3 基于小波压缩域

小波是一种处理多尺度可视化信息的强有力的数学工具,图像的小波表述则给出了图像在不同尺度下的变化信息 (Mallat S. 1989)。小波母函数  $\psi(x)$  是一个振荡衰减的函数 (并且它在  $x$  为无限远处收敛到零), 对于一维信号函数  $f(x)$ , 设  $\psi_{2^j}(x) = 2^j \psi(2^j x)$  是  $\psi(x)$  在尺度  $2^j$  的伸缩小波系, 设  $N$  是  $f(x)$  的采样点数, 则  $f$  在  $2^j$  尺度的细节信号为  $f$  与平移伸缩小波的内积, 即

$$W_{2^j} f(n) = \langle f(u), \psi_{2^j}(u - 2^{-j}n) \rangle, 0 \leq n \leq 2^j N \quad (6-23)$$

可以在不同的尺度  $2^j$  (即在尺度  $1/2, 1/4, \dots, 2^j, j \in \mathbb{Z}$  且  $j \leq -1$ ) 分别利用式 (6-24) 计算其小波系数

$$W_f = (W_{2^j} f)_{-J_{\max} \leq j \leq -1}, J_{\max} = \log_2 N \quad (6-24)$$

接着用一种金字塔算法根据在尺度  $2^{j+1}$  的小波系数来计算尺度  $2^j$  的小波系数, 即

$$A_{2^j} f(n) = \sum_{k=-\infty}^{\infty} h(k-2n) A_{2^{j+1}} f(k), 0 \leq n \leq 2^j N \quad (6-25)$$

$$W_{2^j} f(n) = \sum_{k=-\infty}^{\infty} g(k-2n) A_{2^{j+1}} f(k), 0 \leq n \leq 2^j N \quad (6-26)$$

式中,  $h$  是尺度离散滤波器;  $g$  是小波离散滤波器。

由于小波变换具有良好的时频局部性及与人眼视觉特性相符的多分辨率分析能力, 因此它一经出现就被广泛地应用于图像压缩领域。近几年的研究表明基于小波变换的静止图像压缩系统在性能上优于基于 DCT 的压缩系统, 从而促使小波变换在 JPEG2000 和 MPEG24 中(静态纹理的编码)获得应用。其主要优点可归纳为:

- (1) 多分辨率性能;
- (2) 较好的方向选择性;
- (3) 高去相关性, 高能量集中性;
- (4) 没有块效应;
- (5) 较好的人眼视觉特性。

目前针对小波压缩域的图像检索技术已进行了多方面的研究。一部分学者致力于利用小波系数来进行纹理特征的提取。Smith (Smith J.R. et al 1994) 提出了一种纹理识别的方法, 该方法是首先利用子带能量来定义纹理特征序列, 然后对图像进行 3 级离散小波变换, 并计算所有高频子带小波系数的幅度值, 所有幅度系数子带通过上采样恢复到与原始图像同样的尺寸, 共得到 9 个纹理通道, 所有通道中位置相同的点构成一个 9 维矢量, 对每一矢量中的分量进行二值量化处理, 处理后的矢量共有  $2^9 = 512$  种可能值, 从而可构建 512 级统计直方图, 并用该直方图作为纹理特征来支持图像的检索。虽然这种方法具有较好的检索性, 但是计算复杂度较高。Mandal 等 (Mandal M.K. et al 1999) 在此基础上提出了一种改进算法——快速小波直方图检索方法, 在基本不影响检索性能的情况下, 大大减低了计算复杂度。这一类方法被统称为小波直方图法, 是小波压缩域的典型算法之一, 不但对纹理图像具有较好的检索效果, 而且对于自然图像也很有效。Jacobs 等人基于小波变换系数直接比较实现了一种快速的多分辨率图像检索算法 (Jacobs C.E. et al 1995) 在该算法中所有图像的尺寸都调整为  $128 \times 128$  并进行小波分解, 对每一幅图像选出幅值最大的  $M$  ( $40 \sim 60$ ) 个小波系数记录其颜色、符号以及索引位置, 将这些数据组织起来作为表征图像的特征支持检索。虽然该参考文献给出了较好的检索性能, 但由于特征数据中包含系数的位置索引, 所以该算法明显不具备对平移和旋转等几何变化的鲁棒性 (王晓华 2004)。Wang (Wang J.Z. et al 1997) 等人提出了另一种基于小波分解系数比较的图像检索算法。所有图像也调整为  $128 \times 128$  的尺寸并进行四级小波分解, 利用最低一级的四个  $8 \times 8$  子带图像 (LL, LH, HL, HH) 进行三级检索: 第一级通过 LL 子带系数方差的比较筛选出 20% 的图像; 在第二级中对上一级筛选出的图像进行 LL 子带小波系数的直接比较; 最后一级通过 LL, LH, HL, HH 四个子带图像的系数的比较得到最终的检索输出图像。虽然该文献给出了比 Jacobs 更好的检索性能, 然而该算法同样也不具备对平移、旋转的鲁棒性。

基于小波压缩域的另一种典型方法是子带能量法。这种方法的基本思路是: 对原始图



像做小波分解,得到多个时频子带,计算每个子带的能量形成多维特征矢量用于检索。Chang 提出的通过不规则树分解来进行纹理分析的方法是子带能量法的雏形 (Chang T. et al 1993),在这种方法中通过计算  $J$  个中分辨率子带系数的能量形成  $J$  维特征矢量用于纹理匹配,取得了较好的效果。Lee 等在文献 (Lee M.C. et al 2000) 中为了降低小波包分解时能量特征矢量的维数,仅仅选择包分解子带中能量较大的 7 个子带形成特征矢量用于纹理分类,实验表明在降低计算复杂度的同时,也可以取得很好的分类结果。(Albanesi M.G. et al 2000) 提出了一种利用各子带之间的相关性进行图像检索的方法:首先做小波分解,分别用各级的 LL 子带,LL+HL 子带,LL+LH 子带,LL+HH 子带近似重构上一分辨率图像,再计算它们与用 LL+HL+LH+HH 精确重构的图像之间的相关性,在每一层形成一具有四个分量的特征矢量,用来进行图像分级检索。近年来,很多研究者通过选取不同的小波基(正交基、双正交基)、不同的小波分解方法(塔式小波分解、小波包分解等)以及不同的子带能量计算法等对该类方法进行了深入研究。基于小波子带方面的研究还有文献 (Ma W.Y. et al 1995) 和文献 (Chang S-F et al 1995)。

与前述基于小波子带能量的方法不同,Mandal (Mandal M.K. et al 1996) 提出了一种通过比较方向性子带的直方图来进行纹理匹配的方法。虽然不同图像在整个直方图上可能是相似的,但是不会具有类似的子带统计信息。借助不同级别的水平、垂直、对角信息的不同,可以区别不同图像。在此基础上,Mandal 等人把小波分解的方向子带统计直方图特征用于图像检索 (Mandal M.K. et al 1998),并构造了不同子带的矩特征用于支持小波域的图像检索。同时,由于小波系数的分布近似符合广义高斯分布,Mandal 又提出对不同子带统计特性的匹配只需比较其用于描述分布的标准方差和形状参数,从而实现了一种低复杂度高效率的图像检索算法。Chen (Chen J.L. et al 1994) 则利用小波和隐含的马尔科夫模型提出了旋转和灰度不变的纹理图像分类技术:首先从小波的各子带中提取灰度不变特征,然后利用各子带序列组成的 Markov 模型捕获旋转的变化趋势,这样能做到纹理匹配时的灰度、旋转不变。Wang 和 Mandal 等提出了以小波系数子带的统计特性为索引的技术。这里,每个子带小波系数的一阶矩(均值是其估计)和二阶矩(标准差是其估计)被当作图像索引。这个索引是一个建立在每个小波子带系统上的矩身量。假设子带  $b$  包括  $k_b$  个系数, $\mu_b$  和  $\sigma_b$  分别表示子带  $b$  相关系数的均值和标准差, $\mu_b$  和  $\sigma_b$  如下式:

$$\begin{aligned}\mu_b &= \sum_{i=1}^{k_b} \frac{Wf_i}{k_b} \\ \sigma_b^2 &= \sum_{i=1}^{k_b} \frac{(Wf_i - \mu_b)^2}{k_b}\end{aligned}\quad (6-27)$$

经过  $n$  级小波分解后,小波系数分布在  $3n+1$  个子带中,每个子带都有  $\mu_b$  和  $\sigma_b$  对,总共  $3n+1$  个  $\mu_b$  和  $\sigma_b$  对被用于创建索引。

Sun 等 (Sun M. et al 1995) 证明了任何一种对称小波基都可以作为边缘检测算子来实现图像的边缘检测,只是采用反对称小波基时,检测出的边缘图像会产生半个像素的平移。这为小波压缩域基于边缘轮廓的检索鉴定了理论基础。双正交小波因其具有的良好特性(如线性相位、高阶消失矩等)被广泛应用于图像压缩领域,魏海等 (魏海等 2001b) 依据反对称双正交小波具有的多尺度边缘提取能力,在小波变换域内利用边缘像素的方向梯度相角信息构建直方图统计特征量用来表征图像的内容,提出了一种小波变换域内基于方向梯

度相角直方图的图像检索算法。(魏海等 2002) 证明了基于反对称双正交小波的塔式分解数据能够实现多尺度图像边缘提取。采用一定的反对称双正交小波对图像进行塔式分解, 并通过半重构和局部模极大值检测过程可以得到多个分辨率级上的边缘图像, 每个分辨率级上的边缘图像中的每个边缘像素都具有一个梯度模值和一个方向梯度相角值, 假设第  $j$  级分辨率上的每一边缘像素点所具有模值为

$$M_j(x, y) = \left[ \left( \frac{\partial f_j(x, y)}{\partial x} \right)^2 + \left( \frac{\partial f_j(x, y)}{\partial y} \right)^2 \right]^{1/2} \quad (6-28)$$

方向梯度相角为

$$A_j(x, y) = \tan^{-1} \left[ \frac{\left( \frac{\partial f_j(x, y)}{\partial y} \right)}{\left( \frac{\partial f_j(x, y)}{\partial x} \right)} \right] \quad (6-29)$$

式中,  $f_j(x, y)$  为图像的第  $j$  级近似,  $x$  和  $y$  表示了边缘像素点在模图中的位置。利用每一分辨率级上的边缘图像进行方向梯度相角直方图的构造, 从而来表征图像的内容特性。理论分析和实验结果表明该算法不仅具有较高的检索效率, 同时具有较强的光照变化鲁棒性和一定程度的抗几何变化(尺度、平移、旋转等)的能力。

参考文献(姚玉荣等 2000) 也实现了一种基于小波和不变矩的形状检索方法, 对原始图像进行小波变换, 求取多尺度边界图像; 计算每一边界图像的 7 个不变矩形成特征矢量并归一化; 计算查询图像与目标图像特征矢量之间的距离来确定他们之间的相似性。文中实验表明该方法能较好地描述图像的形状及空间分布信息, 并具有较好的平移、尺度和旋转不变性。

除上述这些直接基于小波变换系数提取各种表征图像内容特征的方法外, 最近不少研究者们开始研究是否能直接从压缩码流中提取表征图像内容的特征。因为小波零树编码方法被公认为是当前最好的基于小波变换的图像编码方法, JPEG2000 中的熵编码部分就采用了零树编码的思想, 所以针对小波零树编码算法研究者提出一些简单有效地检索技术。Wilson 等 (Wilson B. et al 2000) 通过研究小波零树编码算法, 统计 EZW 每一编码层的重要系数数目, 并结合各层的阈值来计算各子带能量。这种方法不需要重建和存储整个小波分解矩阵, 而是边扫描压缩符号流边计算特征矢量, 理论分析和实验结果表明这种方法可以在大大减少计算量和资源需求量的基础上得到较准确的子带能量特征。随着 JPEG2000 压缩标准的推出和普及, 这种基于压缩符号流的检索技术将会得到更广泛的研究。

其他基于小波压缩域方面的研究还有: 利用小波的极坐标表达方法研究其具有的旋转不变性 (Qi F. et al 1994); 研究非线性小波所具有的抗尺度、旋转、平移等几何不变的特性 (Rashkovshiy P. et al 1994); 利用小波系数的局部极大值(模极大值)提取多尺度的边缘信息 (Froment J. et al 1996); 利用小波变换所具有的时频局域性进行特征点检测, 然后基于这些感兴趣的特征点提取纹理、颜色特征实现图像检索 (Loupas E. et al 2000, Sebe N. et al 2000)。

在 JPEG2000 标准没有发布之前, 许多学者在小波压缩域中对图像检索技术进行了大量的探讨。然而, 在 JPEG2000 标准发布以后, 在这个标准框架内探讨图像的检索技术更具有现实意义, 许多学者在这方面已进行了不少的工作。Liu (Liu C. et al 2000) 在 JPEG2000

框架内给出了一种渐近位平面索引技术,它用小波系数的位平面重要位映射图以及重要位直方图作为图像的索引,然后通过索引的匹配进行图像检索。他们还考虑了在压缩码流的包头信息中提取了图像的特征(Liu C. et al 2001),虽然速度加快,但检索效果不好。Bhalod 等人(Bhalod J. et al 2001)在该标准框架内结合感兴趣区域编码的特点给出了基于区域的图像检索方法,它通过 ROI 区域的跟踪和分析,提取 ROI 区域的轮廓以及区域内的纹理和颜色特征作为图像检索的依据。Xiong 等(Xiong Ziyu et al 2002)主要从节省内存的角度提出了一种基于块的图像索引方法,他提取的特征是块内的方差,但提供的信息量较少,效果不明显。Lin(Lin Ni 2003)给出了一种比较新颖的基于树距离比较的 JPEG2000 压缩域图像检索方法,它通过一种名为信息树的结构去描述小波系数的分布,然后通过信息树的距离比较进行图像检索。以上这些相关方法都是在小波压缩域内结合 JPEG2000 压缩特点进行图像检索的,在检索速度上有了很大的提高,但检索效果与传统的方法相比在整体上并没有太大的改进,因此,还需要进一步对该问题进行研究和探讨。

### 6.3.4 基于K-L变换域

K-L 变换是基于图像的统计特性的,其基函数是图像自相关矩阵的特征矢量,它具有最大的能量集中性,是统计最优变换。设给定一组  $N$  个随机矢量,即

$$x = [x_1 \quad x_2 \quad \cdots \quad x_N] \quad (6-30)$$

式中,每个矢量  $x_i$  含  $M$  个分量,即

$$x_i = [x_{1i} \quad x_{2i} \quad \cdots \quad x_{Mi}]^T \quad i=1,2,\cdots,N \quad (6-31)$$

这组随机矢量的均值矢量为

$$m_x = E\{x\} \quad (6-32)$$

式中,  $E\{\bullet\}$  代表期望值,下标  $x$  表示  $m$  所对应的一组随机矢量。这组随机矢量的协方差矩阵为

$$C_x = E\{(x - m_x)(x - m_x)^T\} \quad (6-33)$$

根据矩阵理论,如果矩阵  $C_x$  是一个实对称矩阵,则总可以找到它的一组  $N$  个正交特征值。现令  $e_i$  和  $\lambda_i (i=1,2,\cdots,N)$  分别为  $C_x$  的特征矢量和对应的特征值,并且这些特征值单调排列,即  $\lambda_i \geq \lambda_{i+1} (i=1,2,\cdots,N-1)$ 。再令  $A$  为由  $C_x$  的特征矢量组成其各行的矩阵,并且  $A$  的第 1 行为对应最大特征值的特征矢量, $A$  的最后 1 行为对应最小特征值的特征矢量。如果设  $A$  是将  $x$  转换为  $y$  的变换矩阵,则

$$y = A(x - m_x) \quad (6-34)$$

式(6-34)就称为 K-L 变换,又称为霍特林变换,由这个变换得到的  $y$  矢量的均值为 0,即

$$m_y = 0 \quad (6-35)$$

而且  $y$  矢量的协方差矩阵可由  $A$  和  $C_x$  得到

$$C_y = AC_x A^T \quad (6-36)$$

$C_y$  是一个对角矩阵,它的主对角线上的元素正是  $C_x$  的特征值,即

$$C_y = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{bmatrix} \quad (6-37)$$

它的主对角线以外的元素均为零, 这表明  $y$  矢量的各元素是不相关的, 考虑到  $\lambda_i$  也是  $C_x$  的特征值, 并且沿  $C_x$  对角矩阵的主对角线上的元素是  $C_x$  的特征值, 所以可知  $C_x$  和  $C_y$  具有相同的特征值和相同的特征矢量 (章毓晋 2002)。

从 K-L 变换的定义可以看出, 由于其基函数是自适应的, 因此可以通过映射图像到 K-L 空间, 然后比较 KLT 系数来实现图像检索。该变换具有很好的图像检索能力, 在人脸的识别上得到了较好的效果, 但是由于其固有的复杂性, 其应用受到了很大的制约。

Pentland 等利用 K-L 变换给出了一种人脸识别技术 (Pentland A. et al 1994)。首先, 随机提取一组人脸图像, 根据这组图像计算最优变换矩阵, 然后利用该矩阵计算查询图像和目标图像的 KLT 系数, 最后计算二者 KLT 系数之间的欧式距离, 并以此距离进行人脸识别。因为 KLT 系数只有最前面的系数比较大, 所以可以利用前面的少数系数, 这样就可以降低特征矢量的维数, 减少图像匹配的计算量。虽然 K-L 变换提取的是一幅图像的最主要特征, 但常常携带一些与目标识别无关的信息, 如光照方向、光照强度等, 即使增大样本特征的数量也无法消除这些无关信息。针对这一问题, Swets (Swets D.L. et al 1996) 提出了一种 DKLT (Discriminant K-L Transform) 变换, 主要思路是: 对图像先进行 K-L 变换, 然后对变换的结果进行区别分析, 从而得到一组最具有区别能力的特征。在 DKLT 中, 不同类之间的距离被最大化, 而同类之间的距离被最小化, 其实验结果表明该方法在检索效率上要优于 KLT 方法。但计算量几乎是 KLT 的 2 倍, 另外还需要人工标注每幅训练图像类别, 只能适用于实时性不高的场合, 如人脸识别等, 而对于网上查询、大容量数据库检索等是不适合的。

到目前为止, 由于其固有的计算复杂性, K-L 变换很少被用于传统的图像压缩。然而, K-L 变换在多波段图像的分析处理和编码中应用却比较广泛, 在遥感图像分析与检索方面具有一定的潜能 (李晓华 2004)。

## 6.4 空间域和变换域的融合检索

基于空间域和基于变换域的几种图像检索方法各有优劣。一种检索方法是否能得到普及应用, 很大程度上取决于其所基于的压缩方法的性能和使用范围。KLT 虽然统计最优, 但其固有的计算复杂性阻碍了本身及与之相关的检索技术的广泛应用。DFT 在图像分析中具有重要的作用, 但 DFT 对非周期图像信号压缩能力的差强人意, 很少被用于图像压缩, 其检索能力也就无从谈起。矢量量化的编码过程本身就是一种索引机制, 最易实现压缩域图像检索, 但矢量量化的编/解码过程很不对称。分形虽在编码方面具有很大的潜能, 但是分形码是高度非线性且依赖于待编码图像, 只有图像本身具有明显的自相似性或统计相似性时, 才能获得很高的压缩效率, 另外分形的编/解码过程也很不对称。由于这些原因, 基于 DFT、KLT、矢量量化和分形等的图像检索技术在人们进行了一些探索性研究后, 并未成为压缩域检索技术的研究主流, 不过在一些特殊的领域, 如多波段遥感图像、医学图像,

由于要根据图像自身的特点选取最合适的压缩方法，这些检索技术仍然有其应用潜力。为了充分利用各种技术的优点，一些专家们致力于研究空间域和变换域相融合的图像检索技术，取得了较好的成果。

Idris (Idris F. et al 1995c) 提出了一种在小波域利用矢量量化的图像检索技术，根据小波各子带的特点，对其系数进行矢量量化，并利用其结果进行图像比较。而 Swanson 等 (Swanson M.D. et al 1996) 设计的一种基于分割的图像编码算法的主要思想是：对于所分割的物体对象利用小波、量化的方法编码以提供可直接面向对象的图像索引能力，而对残差图像则进行基本的 JPEG 编码。Podilchuk (Podilchuk C. et al 1996) 在 DCT 压缩域采用矢量量化的方法对 DCT 系数进行量化，并利用矢量量化的索引技术进行人脸识别，其结果令人满意。Swanson (Swanson W.M. et al 1998) 等提出一种基于小波域的 VQ 检索方法，它不同于传统的基于内容检索——将特征矢量单独存储为特征矢量数据库的方法，而是将特征矢量存储为压缩图像的一部分，大大简化了检索复杂度对存储空间的要求。算法首先采用一定分割方法在图像中找到感兴趣物体，而后对物体进行小波变换和嵌入式矢量量化，其他的非物体图像区域使用简单的编码算法如 JPEG、小波或矢量量化编码。如图 6.18 所示。其中，多分辨率头文件采用从低分辨率到高分辨率的组织结构，检索过程利用检索图像的有限 VQ 字典得到的适当码字进行，开始寻找各幅图像最低分辨率下的码字，如果匹配，则说明在该图像中找到在此尺度下和检索图像相似的物体，而后可以在更高分辨率下对码字进行匹配，逐步找到相似图像。

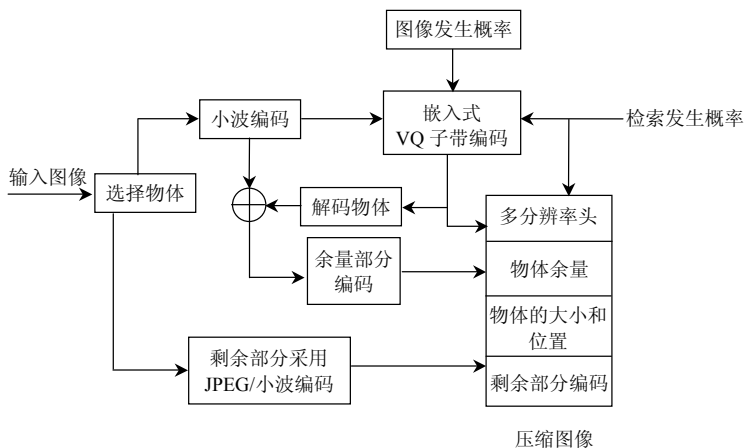


图 6.18 小波域的 VQ 检索方法示意图 (Swanson W.M. et al 1998)

## 6.5 DCT压缩域内的纹理特征

传统的基于 JPEG 图像的检索首先要对压缩图像解压缩，而后利用基于像素域的图像检索方法提取颜色、纹理、形状、草图和空间位置关系等特征，采用一定的相似度度量准则进行检索。这些方法取得了很好的效果，但也存在一些缺点，如它们使用的特征种类比较单一，一般只单独利用纹理或者彩色特征，没有充分利用 DCT 系数所包含的丰富信息，特征的组织形式比较简单，大都采用直方图或随机块的形式，没有充分利用图像的结构性质。

信息。而 DCT 系数有其内在的优良性质，多数统计特征可直接通过图像子块 DCT 直接统计，这些统计特征又直接反映了图像的视觉特性，比在空间域进行类似特征统计分析大大缩短了运行时间。

从图 6.1 和图 6.8 可以看出，基于 JPEG 压缩域的图像检索数据源可以来自 JPEG 解压过程的任意环节，针对不同的数据源格式和具体应用，可以有不同的检索模式。由于熵编码属于非结构化、非字节对齐编码，在熵编码后进行特征提取操作十分困难，所以现有的检索方法通常都在熵解码后或熵解码和反量化后进行特征提取。而且，如果适当组织 DCT 系数，使之反映出图像纹理的方向性，则有利于利用 DCT 系数进行图像检索。本节首先对 JPEG 图像进行熵解码得到量化的 DCT 系数，然后在 DCT 压缩域中，利用 DCT 系数的分布特点，介绍一种图像纹理特征的提取算法——复杂度直方图方法（赵珊等 2007b）。该算法根据数据特点，统计每个 DCT 块的复杂度，构造复杂度直方图来表征原始图像的纹理特征。它不受旋转变换、平移变化等的影响，能够反映图像的基本信息，因此有利于图像检索。同时，考虑到 DCT 系数在块中的空间位置分布的不同对最后检索效果的影响，选取每个 DCT 块中能量最大的九个系数作为重要系数，以此来对每个块的复杂度设置权值，从而避免了由于复杂度相同而系数空间分布信息不同而造成的误检和漏检情况。该方法方便简单，直接在 DCT 域提取图像的特征矢量，计算复杂度低，而且能很好的反映图像中的纹理分布。实验结果表明，该算法具有较好的检索效果。还可以根据 DCT 块各系数所处位置的空间信息改进复杂度直方图，进一步提高图像的检索效率。

### 6.5.1 复杂度的定义

近年来，在生物医学信号处理，特别是在计算机图像分析中，复杂度分析引起了人们广泛的关注（沈恩华等 2000）。复杂度方法是非线性动力系统的一种度量方法，它是作用于时间序列的一种指标。一段时间序列通过一个给定的复杂度算法处理，可以得到一个复杂度值，这个值表示的是序列的非规则程度。不规则变化序列的复杂度值较高，像常数序列、周期序列这种模式比较简单的序列的复杂度值就比较低，这里指的是一维复杂度。参考文献（陈芳等 1998）提出了一种 CO 复杂度的概念，主要思想是：把信号分解成规则成分和非规则成分两部分，CO 复杂度定义为非规则部分在原信号里所占的比例。复旦大学数学系蔡志杰副教授严格证明了 CO 复杂度的一些性质（沈恩华 2004），表明 CO 复杂度一定条件下可以作为时间序列随机程度的指标，在随机性复杂度的意义下也可作为复杂度的一个定量指标。但由于二维空间与一维空间存在着根本的区别，一维的符号序列存在自然、明确的，可以从左或者从右给序列中的每个符号编上序号，而二维的符号矩阵没有自然的顺序，将每个符号排序的方法也不是唯一的，因此，采用一维复杂度的定义无法很好地描述二维空间的数据分布信息。参考文献（沈恩华 2004）提出了一种用于度量图像等二维结构的复杂度算法，定义如下：

设  $\{f(j,k), j=0,1,2,\dots,M-1, k=0,1,2,\dots,N-1\}$  是一个长度为  $M$ ，宽度为  $N$  的平面图像序列， $\{F_{MN}(m,n), m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1\}$  为相应的二维离散 Fourier 变换序列，定义为

$$F_{MN}(m,n) = \frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j,k) e^{-2\pi i \frac{jm}{M}} e^{-2\pi i \frac{kn}{N}}, \quad m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1 \quad (6-38)$$

式中， $i=\sqrt{-1}$  是虚数单位，为了书写方便，记  $W_M = e^{\frac{2\pi i}{M}}$ ,  $W_N = e^{\frac{2\pi i}{N}}$ ，则  $F_{MN}(m,n)$  可以简

写为

$$F_{MN}(m,n) = \frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j,k) W_M^{-jm} W_N^{-kn}, \quad m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1 \quad (6-39)$$

记  $\{F_{MN}(m,n), m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1\}$  的均方值为  $G_{MN}$ , 则

$$G_{MN} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |F_{MN}(m,n)|^2 \quad (6-40)$$

定义

$$\tilde{F}_{MN}(m,n) = \begin{cases} F_{MN}(m,n), & \text{当 } |F_{MN}(m,n)|^2 > G_{MN} \text{ 时} \\ 0, & \text{当 } |F_{MN}(m,n)|^2 \leq G_{MN} \text{ 时} \end{cases} \quad (6-41)$$

对  $\{\tilde{F}_{MN}(m,n), m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1\}$  作离散 Fourier 逆变换, 得到一个新的二维平面图像序列, 即

$$\tilde{f}(j,k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tilde{F}_{MN}(m,n) W_M^{jm} W_N^{kn}, \quad j=0,1,2,\dots,M-1, k=0,1,2,\dots,N-1 \quad (6-42)$$

二维复杂度可定义为

$$C_0 = \frac{\sum_{j=0}^{M-1} \sum_{k=0}^{N-1} |f(j,k) - \tilde{f}(j,k)|^2}{\sum_{j=0}^{M-1} \sum_{k=0}^{N-1} |f(j,k)|^2} \quad (6-43)$$

蔡志杰副教授证明了二维复杂度具有下面的性质 (蔡志杰等 2004):

- ① 如果图像是单一颜色的 (常数二维结构), 它的二维复杂度为 0;
- ② 如果图像在纵向或横向是周期的, 随着图像不断延伸, 二维复杂度趋向于 0;
- ③ 对于任意图像, 它的二维复杂度总介于 0 和 1 之间; 如果图像是随机的, 每个像素的取值服从于独立同分布, 且有有限 4 阶矩, 记其均值为  $Ef = \mu$ , 方差为  $Df = \sigma^2$ , 则当  $M \rightarrow \infty$  或  $N \rightarrow \infty$  时, 二维复杂度以概率 1 收敛于  $\frac{\sigma^2}{\mu^2 + \sigma^2}$ , 特别地, 当  $\mu = 0$  时, 二维复杂度以概率 1 收敛于 1。

从定义可以看出, 二维复杂度利用一个阈值将序列的傅里叶变换分为规则和不规则两部分, 最后计算非规则部分占原信号的比例, 从而通过频谱来确定图像的不规则程度。显然图像越不规则, 越复杂, 二维复杂度越大。参考文献 (沈恩华 2004) 证明, 二维复杂度能很好地反映图像在直观上体现出的规则和非规则程度, 可以作为图像等二维结构非规则程度的度量。

### 6.5.2 复杂度直方图

JPEG 压缩标准中, DCT 变换的结果使得每一个  $8 \times 8$  的 DCT 块都有 1 个 DC 系数和 63 个 AC 系数。而对一个 DCT 块来说, 原始图像中纹理复杂的 DCT 块, 系数分布比较复杂, 中高频系数非零值较多, 而平滑的 DCT 块中, 系数分布较简单, 中高频系数非零值较少。因此, DCT 块的二维复杂度也可以在一定程度上间接反映原始图像块中的纹理信息。因此

在算法中,我们提取每个 DCT 块的复杂度构造复杂度直方图来描述原图像的纹理特征(赵珊等 2007c)。

假设原图像  $I$  的大小为  $M \times N$ , JPEG 压缩图像采用  $8 \times 8$  大小分块,整幅图像共分子块数目为  $\frac{M}{8} \times \frac{N}{8}$ , 用  $C_{mn}$  来标记  $(m, n)$  处子块的复杂度, 则复杂度直方图可定义为

$$H_c(I) = \frac{64}{M \times N} \Pr[C_{mn} = c], \quad c \in [0, t] \quad (6-44)$$

式中,  $m = 0, 1, 2, \dots, \frac{M}{8} - 1, n = 0, 1, 2, \dots, \frac{N}{8} - 1, t$  为复杂度的量化级数。

用上述算法提取的直方图在一定程度上反映了原始图像中的纹理信息, 但复杂度直方图同颜色直方图一样, 仅仅体现了 DCT 系数的统计分布, 而没有包含其空间分布信息。而对于每个块中的 DCT 系数来说, 不同位置的系数反映了原始图像中不同的方向信息, 如果图像块在水平、垂直、对角方向具有明显的边缘特征, 那么其变换后的 DCT 系数将相应的在水平、垂直和对角方向上的值较大 (Abdelmalek A.A. et al 1994)。因此, 如果不考虑 DCT 系数在块中的分布, 仅仅采用复杂度直方图用于检索, 可能会造成误检或漏检。

考虑到这个因素对最终检索结果的影响, 在算法中, 为体现 DCT 系数的空间分布信息, 利用块中能量最大的前几个 AC 系数来为每个块的复杂度引入权函数。利用能量最大的几个 AC 系数的分布来设定权函数, 是因为它们保持了大量的能量和纹理信息, 能从另一方面来体现块的特征。由于在 DCT 块中与 AC 系数对应的频率越高, 量化效应对它的影响就越大, 其带来的误差也越大。所以, 使用少数 AC 系数比使用较多 AC 系数要好, 而且在整幅图像中有小物体平移时, 这种方法也不受影响, 同时也是由于计算量的考虑, 选取了前 9 个能量最大的系数。

如图 6.19 所示, 虽然经过 DCT 变换后, 低频能量都集中在了图像的左上角, 但是前几个能量最大的低频系数并非如图 6.19 (b) 的排列方式, 当任取图 6.19 (a) 中的一个子块做实验, 发现其能量最大的 9 个低频系数的位置如图 6.19 (c) 所示。同时, 对于具有相同或相近复杂度的 DCT 块来说, 考虑到如果块中各分量的位置不同, 在对 DCT 系数进行排序后, 各系数的移动次数也将不同, 因此, 采用排序时各分量移动次数的差异可有效地对 DCT 块进行区分。算法中, 我们采用某种排序方法 (如冒泡排序法), 按如图 6.20 所示的“Z”字形扫描方式将前 32 个系数进行排序 (许多高频系数已经被量化成“0”, 由于计算量的考虑, 故将它们忽略), 得到其中能量最大的 9 个系数, 引入加权函数  $f(x)$

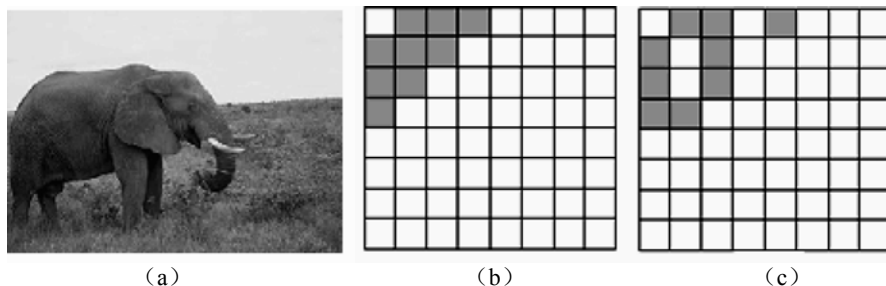


图 6.19 图像 (a) 及其中某个 DCT 块中最大能量系数的分布图 (赵珊等 2007a)



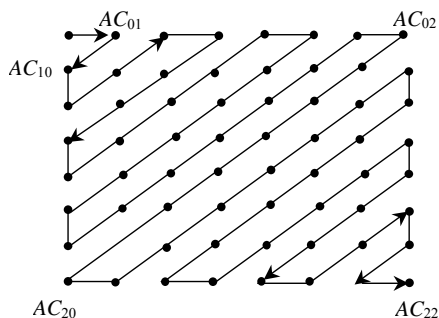


图 6.20 排序所用的扫描方式

$$f(x) = 1 + \frac{m_x}{m_{\max}} \quad (6-45)$$

式中,  $x$  表示任意的 DCT 块,  $m_x$  表示 DCT 块  $x$  经过排序后, 前 9 个能量最大的系数移动的总次数,  $m_{\max}$  表示排序时, 9 个系数需要移动的最大次数, 即对逆序 DCT 系数进行排序时 9 个系数需要移动的总次数。

为了选取合适的量化策略, 我们对复杂度的分布进行了统计, 首先从 Corel 图像库中随机提取 200 幅图像组成测试库, 所选取的图像包括多个类型, 这样可以使统计结果具有一般性, 并假定选取的图像库足够大, 测试库中图像的分布特征足以能描述整个图像集的分布特征, 然后对所有图像的 DCT 块计算其复杂度并进行统计, 以此作为量化的依据。统计结果如图 6.21 所示, 由图可知, DCT 块的复杂度分布不均匀, 大多数复杂度值位于区间  $[0.03, 0.5]$ 。因此, 我们在算法中采用如下非均匀量化的方法来构造复杂度直方图: 在区间  $[0, 0.03]$  上将复杂度值均匀量化为 10 柄, 在区间  $[0.03, 0.5]$  上量化为 120 柄, 在区间  $[0.5, 1]$  量化为 40 柄 (Zhao S. 2008)。

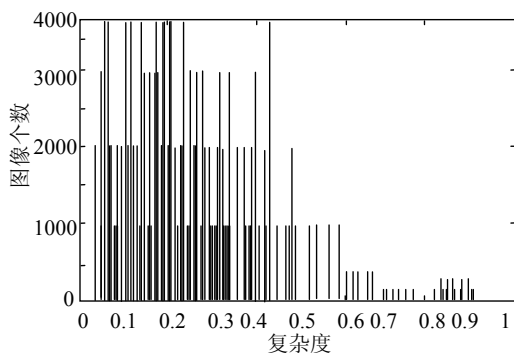


图 6.21 测试库图像 DCT 块分布统计规律

文献 (沈恩华 2004) 已经通过大量的实验证明图像的大小和形状改变对最终复杂度的计算没有影响, 而直方图统计特性通常对图像的旋转和平移等变换具有较好的鲁棒性, 归一化后具有一定的尺度不变性, 因此该算法对原始图像的各种形变具有一定的鲁棒性。算法中通过除图像尺寸大小来进行归一化。

## 6.6 DCT压缩域内的形状特征

上一节讨论的特征仅仅利用了 DCT 系数所隐含的纹理信息,因此,对纹理丰富的图像检索效果较好。然而,对于每一个  $8 \times 8$  的图像块经 DCT 变换后得到的 1 个 DC 系数和 63 个 AC 系数来说,图像的绝大部分能量集中在低频系数中,量化后,大部分高频系数接近于 0,低频系数则含有图像的主要能量。其中 DC 系数代表着块内像素密度均值,低频 AC 系数中含有丰富的图像纹理和边缘信息。因此,为进一步探索 DCT 系数区域所具有的特性,可以利用 DCT 块的系数分布特征来提取图像的边缘及形状信息(赵珊等 2008)。

### 6.6.1 理想边缘模型DCT块的分类

从本章第 3 节 DCT 的定义可以看出,变换后每一个  $8 \times 8$  的 DCT 块中的系数都是块内所有像素值的线性组合。以  $AC_{01}$  为例,即

$$AC_{01} = \frac{C_0 C_1}{4} \left\{ \cos \frac{\pi}{16} \left( \sum_{j=0}^7 f(0, j) - \sum_{j=0}^7 f(7, j) \right) + \cos \frac{3\pi}{16} \left( \sum_{j=0}^7 f(1, j) - \sum_{j=0}^7 f(6, j) \right) + \cos \frac{5\pi}{16} \left( \sum_{j=0}^7 f(2, j) - \sum_{j=0}^7 f(5, j) \right) + \cos \frac{7\pi}{16} \left( \sum_{j=0}^7 f(3, j) - \sum_{j=0}^7 f(4, j) \right) \right\} \quad (6-46)$$

从式(6-46)可以看出,  $AC_{01}$  的值取决于 DCT 变换前  $8 \times 8$  块的左右两部分的亮度值的差,体现了图像在垂直方向上的灰度差,同理,  $AC_{10}$ ,  $AC_{20}$ ,  $AC_{02}$ ,  $AC_{11}$  等几个系数也分别体现了图像在不同方向上分布的灰度差,其物理意义如图 6.22 所示。

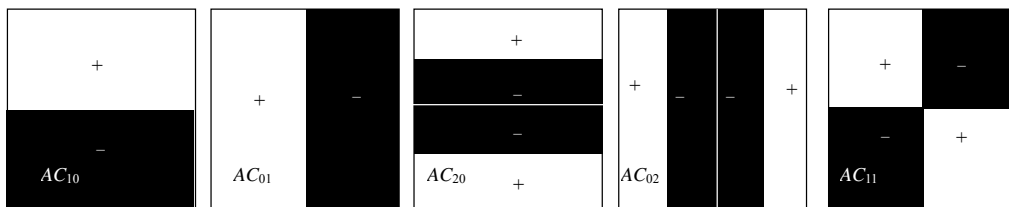


图 6.22 DCT 块中五个 AC 系数的物理意义

在上述分析的基础上,文献(Shen B. et al 1996)根据当只有一条边缘线穿越 DCT 块时可能出现的情况提出了两种理想边缘模型,如图 6.23 所示,  $\theta$  表示边缘方向的角度值,  $d$  表示边缘相对中心的偏移,  $I$  是指边缘强度。对于图像经过 DCT 变换后得到的所有 DCT 块,如果直接在 DCT 压缩域,通过 AC 系数的计算,完全表示出其边缘模型,有两个问题:一方面,要准确地表示出边缘的方向和相对偏移,才能比较准确地表示出边缘,这涉及较多的 AC 系数和很大的计算量;另一方面,要完全表示出边缘的二值模型,需要用 64 个数据单位描述每个块的边缘,这为以后计算增加了负担。针对以上问题,文献(Shen B. et al 1996)提出了一种简化的边缘模型——对于每个  $8 \times 8$  的块,仅仅通过使用前 5 个 AC 系数值的大小,就可以判断出边缘的大概方向,从而可以将其划分为如图 6.24 所示的 20 种边缘类型。

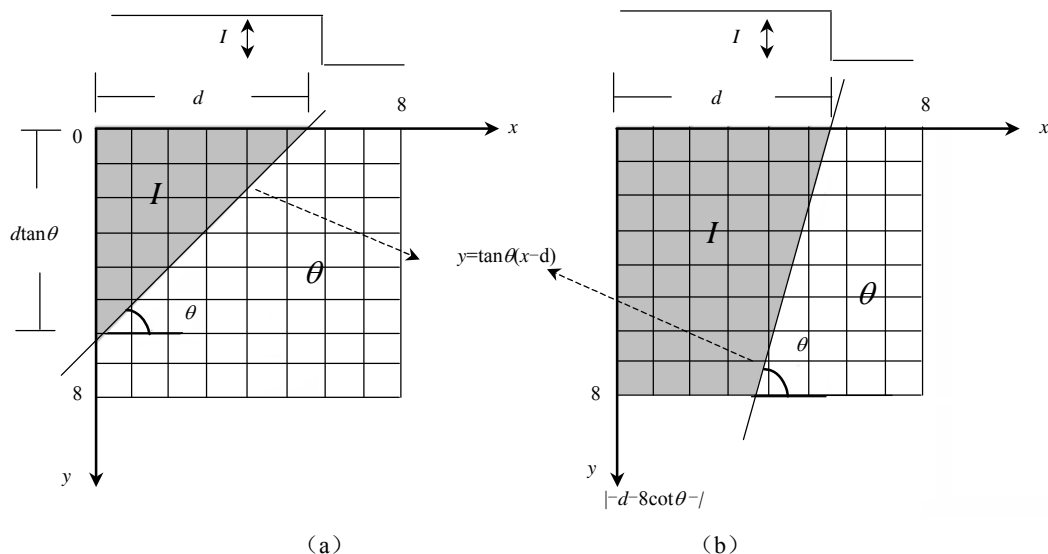


图 6.23 两种理想的边缘模型

$AC_{10} > 0$ $AC_{01} > 0$	$AC_{20}=0$		(b)	$AC_{10} > 0$ $AC_{01} < 0$	$AC_{20}=0$		(b)
	$AC_{02}=0$		(b)		$AC_{20}=0$		(b)
	$AC_{11} > 0$		(a)		$AC_{11} > 0$		(a)
	$AC_{11} < 0$		(a)		$AC_{11} < 0$		(a)
$AC_{10} < 0$ $AC_{01} < 0$	$AC_{20}=0$		(b)	$AC_{10} < 0$ $AC_{01} > 0$	$AC_{20}=0$		(b)
	$AC_{02}=0$		(b)		$AC_{02}=0$		(b)
	$AC_{11} > 0$		(a)		$AC_{11} > 0$		(a)
	$AC_{11} < 0$		(a)		$AC_{11} < 0$		(a)
$AC_{10}=0 \quad AC_{01} > 0$			(b)	$AC_{10} \neq 0 \quad AC_{01} > 0$			(b)
$AC_{10}=0 \quad AC_{01} < 0$			(b)	$AC_{10} \neq 0 \quad AC_{01} < 0$			(b)

图 6.24 利用 DCT 块中 AC 系数之间的关系提取的边缘类型

### 6.6.2 空间边缘分布特征的提取

定义了 DCT 块的边缘类型后，一幅  $M \times N$  的图像  $I$  就对应着一个  $[M/8] \times [N/8]$  的矩阵  $P$ ，其中  $P(x, y)$  的值为  $(x, y)$  处 DCT 块所属的边缘类型索引值。为了提取边缘分布的空间信息，我们针对  $P(x, y)$  中的某一类索引值，保留该索引值的位置上的值，将其他的位置 0，构成一个该类的空间分布图，在此基础上提取边缘的空间分布特征（赵珊等 2007c）。

设  $A_i = \{(x, y) | (x, y) \in P, P(x, y) = i, 1 \leq i \leq 20\}$  表示  $P$  中索引值为  $i$  的所有点的集合。设  $|A_i|$  表示集合  $A_i$  中点的数目， $C_i = (x_i, y_i)$  为  $P$  中索引值为  $i$  的所有点的质心。 $x_i$  和  $y_i$  定义如下：

$$x_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} x; \quad y_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} y \quad (6-47)$$

设  $r_i$  表示  $P$  中  $(x, y)$  处索引值为  $i$  的点同其质心的距离, 其定义如下:

$$r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (6-48)$$

则  $P$  中所有属于  $i$  的点到质心的距离和, 即

$$R_i = \sum r_i = \sum_{(x,y) \in A_i} \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (6-49)$$

算法中, 将每一种边缘类型构成的空间分布图中所有点到其质心的距离和作为表征其空间分布的特征, 从而构造了整个图像的空间分布特征  $(R_1, R_2, R_3, \dots, R_{20})$ 。

由于各个 DCT 块的边缘类型是利用每个 DCT 块的五个系数粗略分类的, 即使相同类型的 DCT 块, 其能量也有可能不同。所以, 对两幅图像而言, 当两者对应边缘类型的空间分布一致时, 由于每类边缘的能量大小不一, 也会造成图像的误检。考虑到这个因素对最后检索效果的影响, 我们以每类边缘空间分布图上值不为 0 的点所代表的 DCT 块的边缘强度总和作为图像的另一个特征。对于图 6.24 中 20 种边缘, 可以利用如图 6.25 所示的对称准则 (Lee S.W. et al 2000) 将其转换为图 6.23 中 (a) 和 (b) 两种理想边缘, 然后计算其边缘强度和。强度的数学定义如下, 具体推导过程见文献 (Lee S.W. et al 2000)。

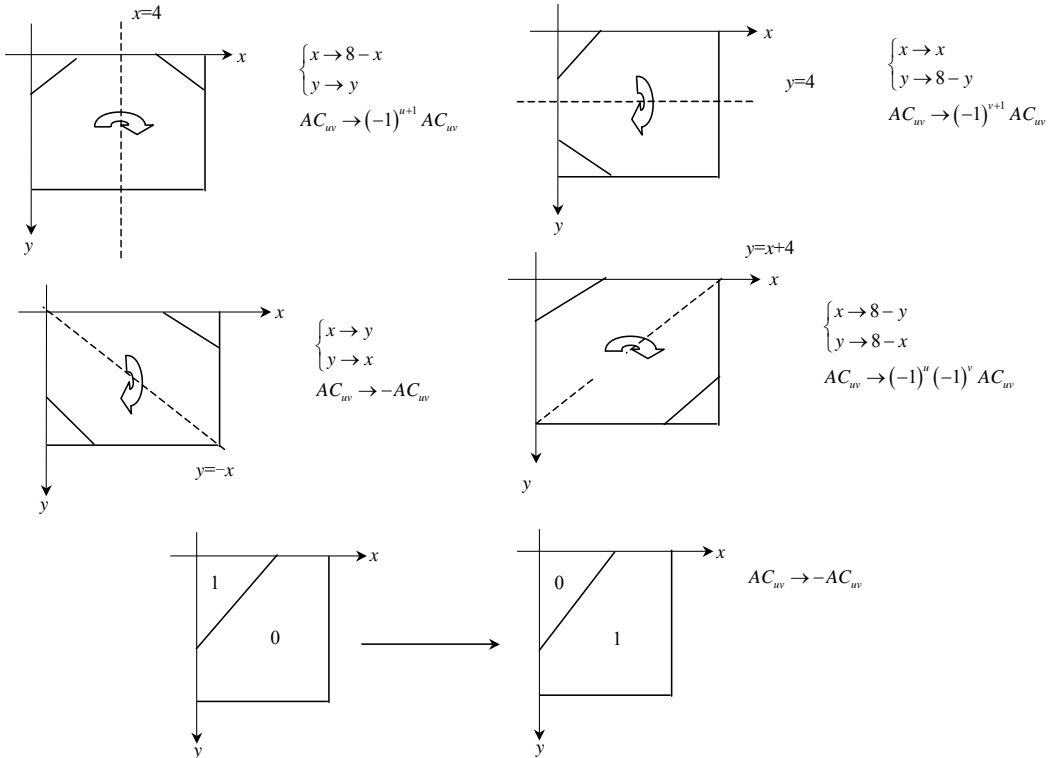


图 6.25 计算 AC 系数的对称准则 (Lee S.W. et al 2000)

$$I_{(a)} = 2\sqrt{2} \left( \frac{\pi}{8} \right)^2 \frac{AC_{10} AC_{01}}{\sqrt{(AC_{10} - AC_{20})(AC_{01} - AC_{02})}} \quad (6-50)$$

$$I_{(b)} = 2\sqrt{2} \left( \frac{\pi}{8} \right)^2 AC_{01} \sqrt{\frac{\sqrt{2}AC_{10}AC_{11}}{\sqrt{2}AC_{10}AC_{11} - AC_{20}AC_{01}}} \quad (6-51)$$

则边缘强度和为

$$E_i = \begin{cases} \sum I_{(a)} & i \in (a) \\ \sum I_{(b)} & i \in (b) \end{cases} \quad (6-52)$$

故提取的边缘强度特征可表示为  $(E_1, E_2, \dots, E_{20})$ 。

## 参 考 文 献

- [1] Abdelmalek A.A., Hershey J.E., 1994. Feature cueing in the discrete cosine domain. *Journal of Electronic Imaging*, 3 (1): 71-80.
- [2] Ahalt S.C., Krishnamurthy A.K., Chen P., et al., 1990. Competitive learning algorithms for vector quantization. *Neural Networks*, 277-290.
- [3] Albanesi M.G., Giacane A., 2000. Fast retrieval on compressed images for internet applications. *Proceedings of the 5th IEEE International Workshop on Computer Architectures for Machine Perception*, Padova, Italy, 136-141.
- [4] Arps R.B., Truong T.K., 1994. Comparison of international standards for lossless still image compression. *Proc.IEEE*, 82 (6): 889-899.
- [5] Augustejin M. L., Clemens E., Shaw K. A., 1995. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. *IEEE Trans Geosci Remote Sensing* 33 (5): 616-626.
- [6] Barbas J.S., Wolk S.I., 1993. Efficient organization of large ship radar databases using wavelets and structured vector quantization. *Proc. of A silomer Conference on Signals, Systems and Computers*, Pacific Grove, CA USA, 1: 491-498.
- [7] Bae H. J., Jung S. H., 1997. Fast texture based on DCT. *Proceeding of ICICS97*, Singapore, 1065-1068.
- [8] Bhalod J., Fahmy G.F., Panchanathan S., 2001. Region based indexing in the JPEG-2000 framework. *Proceeding of SPIE: Internet Multimedia Management Systems II*, 4519: 91-96.
- [9] 蔡志杰, 顾凡及, 沈恩华. Co 复杂度的数学基础. *应用数学与力学*, 2005, 26 (9): 1188-1196.
- [10] Celentano A., Lecce V. D., 1997. A FFT based technique for image signature generation. *Proc SPIE: Storage and Retrieval for image video databases V 3022*, 2, 457-466.
- [11] Chang C.C., Chuang J.C., Hu Y.S., 2004. Retrieval digital images from a JPEG compressed image database. *Image and Vision Computing*, 22: 471-484.
- [12] Chang S-F., 1995. Compressed domain techniques for image/ video indexing and manipulation. *Proceedings of IEEE International Conference on Image Processing*, Washington, DC, USA, 314-317.
- [13] Chang T., Kuo C-C J., 1993. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans image processing*, 10, 2 (4): 429-441.
- [14] Chen J. L., Kundu A., 1994. Rotation and gray scale invariant texture identification using wavelet decomposition and hidden Markov model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (2): 208-214.

- [15] 陈芳, 顾凡及, 徐京华等. 一种新的人脑信息传输复杂性的研究. 生物物理学报, 1998, 14(3): 508-512.
- [16] Chung F.L., Lee T., 1994. Fuzzy competitive learning. Neural Networks, 7 (3): 539-551.
- [17] Climer S., Bhatia S., 2002. Image database indexing using JPEG coefficients. The journal of pattern recognition, Elsevier, 35: 2479-2488.
- [18] Eftekhari-Moghadam A.M., Shanbehzadeh J., Mahmoudi F., et al, 2003. Image retrieval based on index compressed vector quantization. Pattern Recognition. Elsevir, 36, 2635-2647.
- [19] Falconer.K.著, 曾文曲, 刘世耀译. 分形几何-数学基础及其应用. 辽宁: 东北工学院出版社, 1991.
- [20] FanY., Wang R.S., 2002. An image retrieval method using DCT feature. Comput, Sci&Technol., 17 (6), 865-868.
- [21] Feng G.C., Jiang J., 2003. JPEG compressed image retrieval via statistics features. Pattern Recognition, Elsevier, 36 (4): 977-985.
- [22] Froment J., Mallat S., 1996. Second generation image coding and wavelet transform. Proc Int. Conf first world congress of Nolinear analysis, 1923-1932.
- [23] Furht B., Saksobhavivat P., 1998. A Fast Content-Based Video and Image Retrieval Technique Over Communication Channels. Proc. of SPIE Symposium on Multimedia Storage and Archiving Systems, Boston, MA, 11.
- [24] 黄翔宇, 章毓晋. 基于压缩域的图像检索技术研究进展. 中国图像图形学报, 2003, 8 (A) 5, 499-508.
- [25] 黄祥林. 基于压缩域的图像检索技术的初步研究. 博士学位论文, 北京工业大学, 2001.
- [26] 黄祥林, 宋磊, 沈兰荪. 基于 DCT 压缩域的图像检索方法. 电子学报, 2002a, 30 (12): 1786-1789.
- [27] 黄祥林, 沈兰荪. 基于 DCT 压缩域的纹理图像分类. 电子与信息学报, 2002b, 24 (2): 216-221.
- [28] 黄祥林, 沈兰荪. 一种具有旋转不变性的压缩域纹理图像分类方法. 电子与信息学报, 2002c, 24 (11): 1190-1196.
- [29] 黄祥林, 沈兰荪. 基于 DCT 压缩域的图像字符定位. 中国图像图形学报, 2002d, 7 (A), (1): 22-26.
- [30] Huang X.L., Shen L .S., 2002. Texture image classification in DCT compressed domain. Journal of Electronics& Information, 24 (11): 1190-1196.
- [31] Ida T., Sambonsugi Y., 1995. Image segmentation using fractal coding. IEEE Trans Circuits and Systems for video technology 5, 12, 567-570.
- [32] Idris F., Panchanathan S., 1995a. Image indexing using vector quantization. SPIE proceeding: Storage and Retrieval for image and video databases III 2420, 373-380.
- [33] Idris F., Panchanathan S., 1995b. Storage and retrieval of compressed images, IEEE Trans Consumer Electronics, 41 (8), 937-941.
- [34] Idris F., Panchanathan S., 1995c. Image indexing using wavelet vector quantization. SPIE proceedings: digital image storage archiving systems 2606, 10: 269-275.
- [35] Jacobs C.E., Findelstein A., Salesin D.H., 1995. Fast multiresolution image querying. ACM International conference on Computer graphics and interactive techniques, Los Angeles, 277-286.
- [36] Jacquin A.E., 1993. Fractal image coding: a review. Proceeding of IEEE, 81 (10): 1451-1465.
- [37] Lay J. A., Ling G., 1999. Image retrieval based on energy histograms of the low frequency DCT coefficients. Acoustics, Speech and Signal Processing, 6: 3009-3012.
- [38] Lee M-C, Pun C-M, 2000. Texture classification using dominant wavelet packet energy features. Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, Austin, TX, USA,

- 301-304.
- [39] Lee S-W, Kim Y-N, Choi S.W., 2000. Fast scene change detection using direct feature extraction from MPEG compressed videos. *IEEE Transactions on Multimedia*. 2 (4), 240-254.
  - [40] 李鹏杰, 杨树元. DCT 域中 MPEG7 主色描述符的提取. *电子与信息学报*, 2004, 26 (11): 1693-1699.
  - [41] 李晓华, 沈兰荪. 基于压缩域的图像检索技术. *计算机学报*, 2003, 26 (9): 1051-1059.
  - [42] 李晓华. 小波压缩域图像检索技术的初步研究. 博士学位论文, 北京工业大学, 2004.
  - [43] Liu C., Mandal M., 2000. Image indexing in the JPEG2000 framework. *Proceeding of SPIE: Internet Multimedia Management System*, 4210: 272-280.
  - [44] Liu C., Mandal M., 2001. Fast image indexing based on jpeg2000 packet header. *Proceeding of 3<sup>rd</sup> Intl Workshop on Multimedia information retrieval*.
  - [45] Lin Ni, 2003. A novel image retrieval scheme in jpeg2000 compressed domain based on tree distance. *IEEE. The fifth International Conference on Information and Communications Security*, Singapore, 1591-1594.
  - [46] Loupias E., Sebe N., Bres S., et al, 2000. Wavelet2based salient points for image retrieval. *Proceedings of IEEE International Conference on Image Processing*, Vancouver, BC, Canada, 518-521.
  - [47] Ma W.Y., Manjunath B.S., 1995. A comparison of wavelet transform features for texture image annotation. *Proceedings of IEEE International Conference on Image Processing*, Washington, DC, USA, 256-259.
  - [48] Mallat S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on PAMI*. 11 (7): 674-693.
  - [49] Mandal M.K., Aboulnasr T., Panchanathan S., 1999. Fast wavelet histogram techniques for image indexing. *Journal of Computer Vision and Image Understanding*, 75 (1): 99-110.
  - [50] Mandal M.K., Aboulnesr T., Panchanatian S., 1996. Image indexing using moments and wavelets. *IEEE Trans consumer Electronics*, 8, 42 (3): 557-565.
  - [51] Mandal M. K., 1998. Wavelet based coding and indexing of images and video [Ph D dissertation] . University of Ottawa, Ottawa, Canada.
  - [52] Mandelbrot B., 1982. *The fractal geometry of nature*. San Francisco C A, Freeman.
  - [53] 倪林, 苗原. 一种 JPEG2000 压缩域的图像检索方法. *电子与信息学报*, 2005, 27 (3): 474-477.
  - [54] Pentland A., Picard R.W., Sclaroff S., 1994. Photobook: tools for content-based manipulation of image database. *Proc SPIE Storage and Retrieval for image and video database*, 2185: 34-37.
  - [55] Podilchuk C., Zhang X., 1996. Face recognition using DCT-based feature vector. *Proc IEEE Int. Conf on acoustics, Speech and signal processing* 4: 2144-2147.
  - [56] Qi F., Shen D., Quan L., 1994. Wavelet transform based rotation invariant feature extraction in object recognition. *Proc of Int. symposium information theory&its application*, 11: 221-224.
  - [57] Ramamurthi B., Gersho A., 1986. Classified vector quantization of image. *IEEE Transactions on Communications*, 34 (11): 1105-1115.
  - [58] Rashkovshiy P., Sadovnik L., 1994. Scale, rotation and shift invariant wavelet transforms. *Proc SPIE: Optical Pattern Recognition V2237*: 390-401.
  - [59] Reeves R., Kubik K., Osberger W., 1997. Texture characterization of compressed aerial images using DCT coefficients. *SPIE, Storage and retrieval for image and video databases*, 3022: 398-407.
  - [60] Sebe N., Lew M. S., Tian Q., et al, 2000. Color indexing using wavelet-based salient points. *Proceedings*

- of the IEEE Workshop on Content-based Access of Image and Video Libraries, Hilton Head Island, SC, USA, 15-19.
- [61] 邵虹. 基于内容的医学图像检索关键技术研究. 博士学位论文, 东北大学, 2004.
- [62] Shapiro M., 1993. Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans.SP, 41: 3445-3462.
- [63] 沈兰荪. 图像编码与异步传输. 北京: 人民邮电出版社, 1998.
- [64] 沈兰荪, 魏海, 黄祥林. 压缩域图像处理技术研究. 北京工业大学学报, 2000, 26 (3): 24-32.
- [65] Shen B., Sethi I.K., 1996. Direct feature extraction from compressed images. Proc. of SPIE: Storage and Retrieval for Image Video Databases IV, San Jose, CA USA, 2670: 404-414.
- [66] 沈恩华. 脑电的复杂度分析. 博士学位论文, 复旦大学, 2004.
- [67] 沈恩华, 邱志诚, 孟欣等. 脑电图复杂度分析中的粗粒化问题 II. 量化对复杂度计算的影响. 生物物理学报, 2000, 16 (4): 707-710.
- [68] Shneier Michale, Mohamed Abdel-Mottaleb, 1996. Exploiting the JPEG compression scheme for image retrieval. IEEE transaction on pattern analysis and machine intelligence, 18 (8): 849-853.
- [69] Sim D.G., Kim H.K., 2001. Fast texture description and retrieval of DCT-based compressed images. Electronics Letters, 37 (1): 18-19.
- [70] Sloan A. D. 1994. Retrieving database contents by image recognition: New fractal power. Advanced Imaging, 9 (5): 26-30.
- [71] Smith J.R., Chang S. F., 1994. Transform features for texture classification and discrimination in large image databases. Proc IEEE Int. Conf. Image processing 3: 407-411.
- [72] Stone H.S., Li C.S., 1996. Image matching by means of intensity and texture matching in the Fourier domain. Proc SPIE, 2670: 337-349.
- [73] Sun M., Sciabassi R.J., 1995. Symmetric wavelet edge detector of the minimum length. Proceedings of IEEE International Conference on Image Processing, Washington, DC, USA, 177-180.
- [74] Swanson M.D., Hosur S., Tewfik A.H., 1996. Image coding for content-based retrieval. Proc SPIE: VCIP 2727: 4-151.
- [75] Swanson W.M., Tewfik A.H., 1998. Fast progressively refined image retrieval, Journal of Electronic Imaging, 7 (3): 443-452.
- [76] Swets D.L., Weng J., 1996. Using discriminant eigenfeature for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (8): 831-836.
- [77] Vellaikal A., Kuo C-C J., Dao S., 1995. Content-based retrieval of remote sensed images using vector quantization. Proc. Of SPIE: Visual Information Processing IV, San Jose, CA USA, 2488: 178-189.
- [78] Wang J.Z., Wiederhold G., Firschein O., et al, 1997. Wavelet-based image indexing techniques with partial sketch retrieval capability. Proceedings of the 4th International Forum on Research and Technology Advances in Digital Libraries, Washington, DC, USA, 13-24.
- [79] 魏海. 基于小波的压缩域图像检索技术的初步研究. 博士学位论文, 北京工业大学, 2000.
- [80] 魏海, 沈兰荪. 基于分类矢量量化的图像压缩和检索算法. 电子学报, 2001a, 29 (7): 933-936.
- [81] 魏海, 沈兰荪. 一种基于小波分析的迭代分形编码方法. 电路与系统学报, 1998, 3 (4): 82-85.
- [82] 魏海, 沈兰荪, 李晓华. 基于迭代分形的图像压缩和检索方法. 中国图像图形学报, 2002, 7 (A) 11: 1198-1203.



- [83] 魏海, 沈兰荪. 小波变换域内基于方向梯度相角直方图的图像检索算法. 电路与系统学报, 2001b, 6 (2): 20-24.
- [84] 魏海, 沈兰荪. 量化方法及其统计特征量用于图像检索的性能比较. 电路与系统学报, 2001c, 6 (3): 11-15.
- [85] 魏海, 沈兰荪. 反对称双正交小波应用于多尺度边缘提取的研究. 电子学报, 2002, 30 (3): 313-316.
- [86] Wilson B., Bayoumi M.A., 2000. Compressed domain classification of texture images, Proceedings of the 5th IEEE International Workshop on Computer Architectures for Machine Perception, Padova, Italy, 347-355.
- [87] Wohlberg B., Jager G., 1999. A review of the fractal image coding literature. IEEE Trans. IP, 8 (12): 1716-1729.
- [88] Wu D., Tan E.C., 1999. Comparison of lossless image compression algorithms. IEEE TENCON., 6: 718-721.
- [89] 向辉, 石教英. 压缩域多媒体数据处理技术研究. 中国图像图形学报, 1999, 4 (7), 539-543.
- [90] Xiong Ziyoun, Huang Thomas S., 2002. Block-based, Memory-efficient JPEG2000 image indexing in compressed -domain. Fifth IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI): 92-95.
- [91] 杨守义, 罗伟雄. 一种分形域基于内容的图像检索方法. 电子与信息学报, 2003, 25 (3): 419-423.
- [92] 姚玉荣, 章毓晋. 利用小波和矩进行基于形状的图像检索. 中国图像图形学报, 2000, 5 (3): 206-210.
- [93] Yu Hong Heather, 1999. Visual image retrieval on compressed domain with Q-distance. IEEE International conference on computational intelligence and multimedia applications, New Delhi, India, 1013-1016.
- [94] Zhang A., Cheng B., Acharya R. S., 1995. Approach to query by texture in image databases system, Proc SPIE: digital image storage and archiving systems, 2606: 338-349.
- [95] Zhang A., Cheng B., Acharya R.S., et al, 1996. Comparison of wavelet transform and fractal coding in texture-based image retrieval. Proc SPIE: Visual data exploration and analysis III 2666: 116-125.
- [96] 章毓晋. 图像处理和分析基础. 北京: 高等教育出版社, 2002.
- [97] 张问银. 压缩域图像检索及汉字数学表达式研究. 博士学位论文, 中国科学院成都计算机应用研究所, 2005.
- [98] 张问银, 曾振柄, 孙星明. 基于 A/D 能量直方图的 JPEG 图像检索. 计算工程, 2004, 16 (30): 21-22.
- [99] 张问银, 吴尽昭. 一种快速 JPEG 图像检索方法. 计算机工程, 2005, 31 (10): 148-149.
- [100] 赵德斌, 陈耀强, 高文. 基于块方向预测和 Context 的图像无失真编码方法. 软件学报, 1998, 9 (10): 765-769.
- [101] 赵珊. 基于内容的图像检索关键技术研究. 博士学位论文, 西安电子科技大学, 2007a.
- [102] 赵珊, 周利华. DCT 压缩域中基于纹理和形状的图像检索算法. 西安电子科技大学学报, 2007b, 34 (3): 402-408.
- [103] 赵珊, 周利华. DCT 压缩域的图像检索. 北京邮电大学学报, 2007c, 30 (6): 107-110.
- [104] 赵珊, 刘静. DCT 压缩域的图像检索技术. 北京邮电大学学报, 2008, 31 (5): 5-8.
- [105] Zhao S., 2008. Image retrieval based on edge in DCT compressed domain. ITES, VI: 1104-1108.
- [106] Zhong Y., Zhang H.J., et al, 2000. Automatic caption localization in compressed video. IEEE transaction on pattern analysis and machine intelligence, 22 (4): 385-392.

## 第7章 高维索引技术

由于表示图像特征的向量维数较高, 传统的多维索引技术在特征维数足够高的情况下, 近邻搜索性能会低于原始的顺序查找算法, 这种现象被称为“维数灾难”。随着图像库容量的迅速增大, 图像检索的实时性要求也日益突出。在传统多维索引技术的基础上探索更加有效的高维索引技术就成为一个急需解决的问题。本章在介绍高维索引技术及发展趋势的基础上, 详细讨论了基于矢量量化的向量近似方法、二次式距离上基于 SVD 的高维索引方法及基于多分辨率分析的高维索引方法。

### 7.1 高维索引技术

索引技术是加速图像相似性检索的关键技术之一, 也是多媒体和数据库领域研究的热点问题。目前, 有关对高维数据检索结构的研究非常多, 如 (Lu G.J. 2002, Yu C. 2002, 崔江涛 2005, 董道国等 2005, 庄毅等 2007) 等。著名杂志 ACM Computing Surveys 刊登了多篇有关多维数据索引技术的综述文章 (Berchtold S. et al 1998b, Gaede V. et al 1998, Bohm C. et al 2001, Chavez E. et al 2001), 国内也有类似的综述文章 (董道国等 2002, 刘芳洁等 2003, 周项敏等 2002, 张明波等 2005, 贺玲等 2006, 于林森等 2007) 等。在研究者提出的这些索引结构中, 每种结构都有自己性能比较突出的适用范围, 同时数据的类型、分布的情况、维数的大小等都会对索引结构的性能产生很大的影响。

在多维索引结构中, 树型索引结构最为常用, 它的构建方法实际上是基于数据空间的层次化聚类原则。从结构上来讲, 多维树型索引结构类似于一维数据索引结构—B 树 (严蔚敏等 1992)。它的结点分为叶子 (数据) 结点和目录结点, 向量存储在数据结点中, 空间位置临近的向量尽可能存储在同一结点。结点之间以层次化目录结构来组织, 每一个目录结点都指向下一级的一个子树。通常, 数据结点的数据结构与目录结点的结构完全不同。比较而言, 目录结点在各个索引级上具有更好的一致性。在这种层次化组织结构中, 一般存在一个被称之为根结点的单独目录结点, 它是进行查询和更新处理的开始结点。目前的索引结构大都采用平衡树, 即对任何一个目录结点, 其所有子树的高度 (深度) 差不超过 1。树型索引结构的查询性能主要与树型结构的高度有关。

按照结点的建立方式不同, 树型多维索引结构可以进一步划分为数据点划分和空间划分两大类。基于数据点划分的索引结构包括 R 树 (Guttman A. 1984)、R+树 (Sellis T. et al 1987)、R\*树 (Beckmann N. et al 1990)、X 树 (Berchtold S. et al 1996)、SS 树 (White D.A. et al 1996)、SR 树 (Katayama N. et al 1997) 等。在这些索引结构中, R 树, R+树, R\*树, X 树的目录结点采用 MBR (Minimum Bounding Rectangle, 最小外接矩形) 表示, 一个结点的最小外接矩形包络 MBR 定义为包围该结点的所有子树中的点的最小边界矩形。SS 树的目录结点采用超球 (sphere) 表示, 一个结点的超球体包络定义为包围该结点的所有子树中的点的超球体, 有关超球包络的计算可参阅 (White D.A. et al 1996)。SR 树则将这两

种目录结点的表示方法相结合。在搜索过程时,根据 MBR 或外接超球可以求出目录结点中所有数据点与查询点之间的距离下界,利用此特性可以决定需要访问的路径。基于空间划分的索引方法则是对多维空间进行重复分割,分割成的不相连的子空间中的数据用结点来表示。空间划分的树型结构有 k-d 树 (Bentley J.L. 1975), k-d-B 树 (Robinson J.T. 1987) 以及四叉树 (Finkel R. et al 1974) 等。

除上述索引结构外,还有 TV 树 (Lin K. et al 1995), 空间填充曲线 (Sagan H. 1994), 金字塔树 (Berchtold S. et al 1998) 等。最近几年,国内的研究人员也提出多种针对不同应用的多种索引结构 (汪祖媛等 2004, 杨志荣等 2003, 冯玉才等 2002, 周项敏等 2004)。

上述多维索引方法在特征维数较少的情况下具有很好的检索性能,但后来的研究和实验却发现,这些索引方法在维数足够高的情况下,检索性能下降很快,会退化到顺序查找方法,导致出现“维数灾难”现象 (Berchtold S. et al 1998, Weber R. et al 1998, 崔江涛 2005)。在基于内容的图像检索系统中,为了达到较好的查询效果,用于表示图像特征的向量往往具有高维特性。目前许多现有的 CBIR 实验原型系统仅仅处理几百或几千幅图像,因此即使对所有图像按序排列并计算相似度也不会带来性能上的显著下降。但是,当图像集的规模变得越来越大时,检索速度就会成为检索系统中的性能瓶颈,就需要采用高维数据索引方法来加速检索过程。

## 7.2 高维索引技术发展趋势

本节介绍高维索引领域的研究热点和发展趋势。首先,介绍目前公认的一类能够适用于高维情况的索引技术——向量近似方法 (Weber R. et al 1998, Ferhatosmanoglu H. et al 2000, Wu P. et al 2001, Cha G. H. 2002a, Cha G.H. et al 2002b, Berchtold S. et al 2000, 叶航军等 2004, 董道国等 2005, Weber R. et al 2000, Tuncel E. et al 2002); 其次,介绍两类有效提升高维索引性能的方法,近似近邻搜索方法和并行索引方法。

### 7.2.1 向量近似方法

研究者对高维数据的空间特性进行了深入的研究,分析了传统多维检索结构在高维空间中的缺陷,提出了一些针对高维索引的查询复杂度模型 (Bohm C. et al 2001, Berchtold S. et al 1997b)。Weber 等人对这些模型加以改进,提出了一种理想状态下的高维索引代价模型 (Weber R. et al 1998)。该模型指出,在向量空间维数足够高的情况下,所有基于数据分割和空间分割的树型索引结构的性能都将低于最原始的顺序查找方法。在此基础上,Weber 提出一种新的基于向量近似 (Vector Approximation file, VA-file) 的索引方法 (Weber R. et al 1998)。VA-file 方法对数据集采用顺序访问方式,不过并不是直接访问原始特征向量,而是访问特征向量的近似表示——近似向量。根据近似向量可以计算出特征向量到查询向量距离的上界和下界,在顺序访问近似向量时,可以通过距离上下界过滤掉大部分数据,只需访问极少一部分原始向量即可完成搜索过程。由于近似向量的存储空间远小于原始数据,而且采用了顺序访问的方式,VA-file 方法可以减少搜索过程中的最耗时的 I/O 访问时间。

在精确索引方法中,基于近似向量的索引方法是目前公认的较好的能应用于高维情况的索引方法。在高维情况下,其检索性能仍优于顺序查找算法,而其他索引结构不能保证

在维数足够高的情况下,检索性能仍优于顺序查找方法。在经典的综述文献(Bohm C. et al 2001)中,将向量近似方法看做一种充满希望的高维索引方法。在著名刊物 IEEE Transaction on Multimedia 上,也相继刊登了多篇向量近似方法的相关文章(Cha G. H. 2002a, Cha G. H. et al 2002b)。

在近似向量思想的基础上,研究者对 VA-file 方法加以改进,提出许多相关索引方法。VA-file 方法中对每一维数据进行均匀分割,生成近似向量。针对非均匀分布的数据集, Ferhatosmanoglu 等人提出  $VA^+$ -file 方法(Ferhatosmanoglu H. et al 2000),该方法首先采用 Karhunen-Loeve 变换(简称 KL 变换)来消除各分量之间的相关性,然后根据变换后各分量的能量做非均匀量化位数分配。在保证分配的总位数不变的前提下,方差最大的分量会得到更多的量化位数,对于均匀分布的数据集, $VA^+$ -file 方法检索性能要高于 VA-file 方法。Wu 等人也对数据分布的不均匀性进行深入研究(Wu P. et al 2001),针对数据分布的不均匀性,在各维分量上分别采用混合高斯模型(Gaussian mixture model)来拟合数据的边缘分布,然后根据边缘概率分布模型对各维数据进行非均匀分割。高斯混合模型对边缘概率分布的描述要比 VA-file 方法中的均匀分布假设更为精确,也能适应各种分布的数据集。Cha 等人提出基于向量近似的 LPC 文件(Local Polar Coordinate file, LPC-file)方法(Cha G. H. et al 2002a),该方法在近似向量中加入局部极坐标(角度)信息来提高近似向量的过滤能力,局部极坐标的引入加强了近似向量对原始向量的近似能力,提高了近似向量的过滤能力,从而提高了检索性能。叶航军等人引入了分类分裂矢量量化方法,将数据集划分为几个聚类,采用  $VA^+$ -file 中的向量建立方法在每个聚类内分别建立近似向量文件(叶航军等 2004)。

除了对近似向量的构建方式加以改进外,研究者还对近似向量的组织结构进行改进。基于向量近似的索引方法中,需要顺序存储近似向量,查询时也需要顺序访问全部的近似向量,因此可以将其他索引结构与近似向量方法结合起来减少近似向量的访问数量。

Berchtold 等人将 MBR 的概念应用于 VA 方法中,提出 IQ 树(independent quantization tree, IQ-tree)的检索方法(Berchtold S. et al 2000)。与 R 树结构不同的是, IQ 树只是建立单层的 MBR 结构,并且根据 MBR 中数据点的分布密度来分配近似位数,在检索时可以根据查询向量与 MBR 之间的距离下界过滤掉不需访问的近似向量。Cha 等人针对非均匀分布的图像数据特征库,提出 GC 树(grid cell tree, GC-tree)的检索结构(Cha G. H. et al 2002b),它是一种依据数据点分布情况建立非平衡的树型结构。首先将数据空间均匀分割成多个超立方体,根据超立方体内数据点的分布密度不同,分割空间的超立方体被划分为聚类结点(Cluster Node)和外部结点(Outlier Node)两类,对于聚类结点,将分配额外的近似位数进行继续分割,聚类结点采用超立方体表示,而外部结点则直接采用原始向量表示。董道国等人提出了 VAR 树方法(董道国等 2005),直接采用 R 树组织近似向量,表示目录结点区域的 MBR 也采用近似向量表示。

目前 CBIR 系统大多采用了相关反馈技术,研究者也提出了许多向量近似方法在相关反馈中的应用方法。相关反馈技术中常采用的相似度度量方法包括二次式距离的方法(Rui Y. et al 1998, Ishikawa Y. et al 1998)和核函数方法(Chen Y. et al 2001, Heisterkamp D. et al 2001)。向量近似方法可以直接支持二次式距离的度量方法,Peng 等人则将向量近似方法应用到核函数方法中(Peng J. et al 2002, Heisterkamp D. R. et al 2003)。

对向量近似方法中的  $k$  近邻搜索算法进行改进,可以将其应用到近似近邻搜索中,在

下一节中将对对其进行介绍。

### 7.2.2 近似检索方法

由于图像特征向量的高维特性,绝大多数的索引结构会受到维数灾难现象的影响,有必要引入一种新的方法或者思路来解决这一问题。在查询过程中采用近似近邻查询,可以使用户在查询结果的精确性和运行时间上取得折中。从索引技术角度来看,传统的索引结构和查询技术都是为了得到精确的查询结果。在给定特征向量和距离度量函数的前提下,获得精确的相似性查询结果。在CBIR系统中,特征向量和度量距离的选取都带有主观性或者试探性,特征向量本身就只是多媒体内容的近似表示,而非精确表示,所以精确近邻搜索并不具有精确意义。研究者针对多媒体信息检索,提出了近似查询的概念,并且提出了许多支持近似查询的索引结构。

近似相似性查询是指在一定的误差允许下,对数据集进行相似性查询,获得相似性查询结果。对于 $k$ 近邻搜索,近似近邻搜索是解决维数灾难问题的一种有效手段(Ferhatosmanoglu H. et al 2001, Arya S. et al 1998, Ciaccia P. et al 2000, Pramanik S. et al 1999),通过在搜索结果中引入一定的错误率,降低搜索时间,提高检索效率。特征向量和度量距离选取的试探性,以及用户判断的主观性,保证了近似近邻搜索是一种符合应用需求的检索机制。目前,近似近邻查询技术是高维数据索引的一个新的研究趋势,采用近似近邻查询,可以满足大多数的实际应用需要(Indyk P. et al 1998)。

与精确索引方法相类似,近似近邻搜索方法一般从以下两个方面来提高查询效率(Ferhatosmanoglu H. et al 2001):

#### (1) 减少数据访问量

减少搜索过程中的数据访问量,既能降低查询过程中的I/O次数,又能降低CPU运算代价,从而提高查询性能。减少数据访问量可以通过三种途径来实现。

第一种是在搜索过程中,根据用户对查询结果的满意程度判断是否终止查询过程。一般说来,对大型图像数据库,不必对数据集中所有的数据进行搜索就可以结束查询过程。该方法既可以根据用户的主观判断,也可以使用一定的终止判别条件来决定是否终止查询过程。当进一步的搜索过程不能显著提升查询结果的正确性时,可以终止搜索过程。

第二种是基于聚类的近似查询。利用聚类算法将数据集分为几个聚类,首先计算查询向量所属某个聚类的概率,在搜索过程中只对概率较高的所属聚类进行查询。在精确索引技术中也可以采用聚类思想,对每个聚类分别建立索引结构,可以一定程度上提高检索性能(Chakrabarti K. et al 2000, 叶航军等 2004),但是在搜索过程中必须对所有的聚类进行访问。在近似查询过程中,只访问少数几个聚类就可完成查询过程,可以大幅度提高查询效率。比较典型的基于聚类的近似查询方法是K-L变换域聚类方法(Ferhatosmanoglu H. et al 2001)。

第三种是针对树型索引结构的,其剪枝目录结点的思想与传统的树型索引结构相同,也需要一定的判别函数来判断是否对特定的结点进行访问。与传统树型结构相比,判别函数通常加强了对结点的剪枝条件。当维数较高时,树型索引结构往往需要访问大部分的数据结点,加强剪枝条件后,可以降低结点访问数量,从而提升查询性能。可以根据用户提供的近似参数调整查询过程,在查询精度和查询速度两方面取得折中。这类索引方法中,比较典型的是BBD树方法(Arya S. et al 1998),SS树方法(Pramanik S. et al 1999)以及

PAC 方法 (Ciaccia P. et al 2000) 等。BBD 树和 PAC 方法都属于  $\varepsilon$  近似搜索, 在搜索过程中放松了对最近邻距离的限制, 引入了距离的错误率  $\varepsilon$ , 保证返回的结果与查询向量的距离不会超过正确结果的  $(1+\varepsilon)$  倍。PAC 方法进一步加强了近似能力, 当返回结果不是小于  $(1+\varepsilon)$  结果概率的某一阈值时, 就可以终止搜索过程。SS 树中采用超球表示目录结点, 作者在分析高维超球中数据分布特性的基础上, 根据查询区域与超球的角度信息来判断是否访结点。

### (2) 降低特征向量的表示长度

降低特征向量的表示长度 (存储容量) 可以通过两种途径来实现。第一种是降低向量的表示维数, 第二种是降低每维分量的表示长度。

第一种方法是用较低维数的向量来近似表示原向量, 由于需要计算的维数降低, 该方法既可以降低搜索过程中的 I/O 代价, 又可以降低 CPU 运算代价。对于欧氏距离, 正交变换是常用的降维方法。在能量集中特性方面, K-L 变换 (又称为主分量分析) 是一种最优的正交变换方式 (Fukunaga K. 1990)。但是, K-L 变换也存在一些难以克服的缺点, 包括运算复杂度很高及不能适应数据更新等。Kanth 等人提出一种近似的 K-L 变换方法 (Kanth K.V.R. et al 1998), 将原特征向量用较低维数来表示, 用 R 树或 SS 树来组织这些低维数据。若数据集中数据分布发生明显变化, 降低查询精度时, 利用原有的索引结构来重新进行 KL 变换, 该方法可以降低 K-L 变换的运算复杂度。Chakrabarti 等人提出一种局部降维 (local dimensionality reduction) 的方法 (Chakrabarti K. et al 2000), 通过聚类算法, 得出数据集中的多个聚类, 然后对每一个聚类采用局部相关性分析的方法, 降低此聚类中的维数。LDR 方法在查询精度和查询效率方面都优于全局降维方法。此外, Egceioglu 等人提出了一种新颖的用于近似搜索的降维方法 (Egceioglu O. et al 2000, Egceioglu O. et al 2004), 用一种内积的近似计算公式, 通过向量之间的内积来计算近似距离。

第二种方法是降低特征向量中每维分量的表示长度, 该方法就是向量近似方法在近似查询中的应用。Weber 等人首先针对 VA-file 方法提出了近似查询算法 (Weber R. et al 2000), 在近似查询过程中, 省略掉第二阶段的搜索过程, 即只对近似向量进行操作, 采用查询向量与近似向量之间的下界作为近似距离。这种近似查询思想也可应用于 VA+-file 方法。Tuncel 等人提出了基于矢量量化的近似查询方法 (Tuncel E. et al 2002)。向量近似方法本质上是一种标量量化, 而矢量量化是标量量化的自然发展, 它是一种高效的数据压缩技术, 能够有效的提高压缩比。与标量量化相比, 矢量量化能够达到更高的压缩比和更小的量化误差, 更适合进行近似近邻搜索。

## 7.2.3 并行索引方法

并行索引技术是提高检索性能的有效方法之一。这里所说的并行索引主要是指将数据按照一定的划分规则分配到多台计算机中分别建立索引结构。各个计算机之间的 CPU 和磁盘操作都是相互独立并行进行的。通过并行处理, 原来对大规模数据的查询就转变为对各个划分子集的并行查询, 并以各个子集查询的最长耗时作为查询时间, 显然这样的查询实际比原来的查询时间要明显低的多。并行处理的方法除了提高查询速度外, 有时由于数据集过大, 磁盘容量有限, 也只能采取分布式并行查询的方式。

对于高维数据的并行检索, 关键是如何在多个磁盘中分配数据。对于树型索引结构,

这也称为反聚类问题。理论上,应该将数据在各个磁盘上分配的更加“稀疏”一些(Berchtold S. et al 1997a),也就是尽量把原来密集的数据分散开。这样,在磁盘上建立索引时,就能降低树型结构中结点重叠率较高的情况,降低查询时间。

为了实现上述目的,有许多相应的并行方法提出。其中最简单的就是 round robin 方法(Du H.C. et al 1982),一种类似于轮盘赌的方法。通过对向量下标进行取模运算,运算结果相同的数据存放在同一个磁盘中。round robin 方法是一种简单有效的并行处理算法。与单一处理器相比较,查询速度明显提升,但是这种不考虑数据的实际分布情况,仅仅简单的摊派分配的方法并不合理。后来许多算法都在 round robin 方法的基础上进行了改进,它们的一般做法是先对数据空间进行划分,然后再分配磁盘,这样就能在空间信息的基础上更为合理地反聚类分配磁盘。当向量空间维数较高时,反聚类算法也会受到维数灾难现象的影响(Bohm C. et al 2001)。

Du 和 Sobolewski 提出了一种磁盘取模(DM)算法(Du H.C. et al 1982):首先将数据空间划分成等间距的胞腔,每个胞腔都是空间中包含一定数量元素的超矩形,然后再将每个胞腔映射到磁盘。Kim 和 Pramanik 在 DM 算法的基础上略加改进,利用了异或运算实现新的映射,称为 FX 方法(Kim M.H. et al 1988)。Faloutsos 和 Bhagwat 将 Hilbert 曲线引入到并行处理算法中,利用 Hilbert 曲线量化的聚类性质将高维空间的数据映射到一维上并对磁盘数取模来实现数据的分配,称为 HI 算法(Faloutsos C. et al 1993)。与 round robin 方法、DM 算法以及 FX 算法比较,HI 算法在近邻搜索有很大优势,因此在并行处理中被经常采用。

Papadopoulos 等人针对一个处理器和多个磁盘的情况提出了两种并行检索算法(Papadopoulos A. et al 1998)。其基本思想与反聚类相类似,当索引结构中分裂产生一个新结点时,新结点的 MBR 将与其兄弟结点做比较,并且与其距离最远的兄弟结点存储在同一个磁盘中。Dantong Yu 在其博士论文中针对 RAID (redundant array of inexpensive disks) 系统也提出了类似的并行索引方法(Dantong Yu 2001)。

### 7.3 向量空间中的高维特性

真实图像特征库中,数据点的分布是非均匀分布的。对于非均匀分布的数据空间,很难对数据特性进行定量描述,不过我们仍可以从均匀分布的数据集中探讨高维空间中的几何特性和统计特性,这些特性对于非均匀分布的数据集仍然具有指导作用(崔江涛 2005)。

#### (1) 体积

高维空间中最明显的效应就是体积的指数效应,一个超矩形或超球的体积将随着的维数的增加呈指数增加(如果边界长度小于 1,将呈指数下降趋势)。

#### (2) 距离

假定一个数据点均匀分布的高维向量空间  $\Omega$  (下文提到的向量空间均采样该定义),其中,  $\Omega = [0,1]^n$ ,  $n$  表示空间维数,  $N$  代表数据集大小。若采用欧氏距离,那么数据点之间的最小近邻距离与维数和数据集大小的关系可以表达为(Berchtold S. et al 2001)

$$NNd(N,n) = \left( \frac{\Gamma(n/2+1)}{N\pi^{n/2}} \right)^{1/n} \quad (7-1)$$

这里  $\Gamma(n)$  代表伽马函数。式 (7-1) 表明, 最近邻距离将随着维数的增加呈线性增加趋势, 但受数据集大小影响较弱, 随着  $N$  的增长而仅仅减小  $N^{1/n}$ 。文献 (Feredman J.H. 1994) 中对于  $n$  维单位超球内的最近邻距离也得出了类似的表达式。

随着维数的增加, 高维空间中的欧氏距离还受到 Concentration 现象的影响, 即在高维空间中, 任意两个向量之间具有高度的相似性 (Francois D. et al 2004)。对于任意两个数据点  $x$  和  $y_i$  之间的欧氏距离, 可以证明 (Beyer K. S. et al 1999)

$$\lim_{d \rightarrow \infty} \frac{\max_i (d(x, y_i)) - \min_i (d(x, y_i))}{\min_i (d(x, y_i))} = 0 \quad (7-2)$$

### (3) 角度

在一个包含多个数据点的超球内, 随着维数升高, 一个超球包含的数据点与球心的距离趋于相同, 大部分数据点都分布在超球的边缘上。如果一个超球包含多个子超球, 随着维数升高, 子超球的半径将与父超球的半径趋于相同, 各个球心也趋于相近。给定一个数据点集合组成的超球, 随着维数增加, 以球心为参考点, 任意两个点之间的角度将趋近于 90 度 (Li J.H. 2001)。

## 7.4 维数灾难现象

### 7.4.1 查询代价模型

设  $q$  代表任一查询向量,  $NN(q)$  为查询向量  $q$  的最近邻, 若采用欧氏距离, 则有如下定义

**定义 1:** 最近邻距离  $NN^d(q)$  定义为

$$NN^d(q) = L_2(q, NN(q)) \quad (7-3)$$

**定义 2:** 最近邻查询超球  $NN^w(q)$  是以  $q$  为球心,  $NN^d(q)$  为半径的超球。

**定理 1:** 假设数据点均匀分布,  $sp^n(q, r)$  是以  $q$  为球心、以  $r$  为半径的  $n$  维超球, 那么  $sp^n(q, r)$  包含数据点  $p$  的概率等于  $sp^n(q, r)$  与向量空间  $\Omega$  相交部分的体积 (Berchtold S. et al 1997b)。

最近邻查询中, 结点表示区域 (MBR 或超球) 到查询点的距离将被事先计算或估计。表示区域的一个很重要的特性就是查询点到结点的距离不大于到其子结点的距离, 也不大于到其子结点中所有数据点的距离, 这通常称为距离下界属性 (Lower Bounding Property)。以 MBR 表示区域为例, 在最近邻查询过程中, 所有与  $NN^w(q)$  相交的 MBR 都将被访问到, 访问过程可以采用 HS 算法 (Roussopolos N. et al 1995), 按距离下界升序访问各个 MBR, 如果某个 MBR 的距离下界大于目前的最近邻距离, 结束查询过程。近邻查询中另外一种常用的搜索算法是 RKV 算法 (Hjaltason G. et al 1995), 它是一种类似于深度优先搜索的算法。

根据文献 (Berchtold S. et al 1997b) 中提出的查询代价模型, 最近邻超球查询可以转换为点查询。最近邻搜索过程中的数据页面访问量等于与查询超球相交的 MBR 的数量。在搜索过程中最近邻距离是变化的, 查询超球半径取其期望值  $E(NN^d)$ 。我们可以估算出



每个 MBR 被访问的概率: 设 MBR 的扩展 Minkowski 和区域用  $M\text{Sum}(mbr_i, E[NN^d])$  表示, 该区域边界上的点与 MBR 边界上的点的最小距离为  $E(NN^d)$ , 如图 7.1 (b) 所示。假定查询点从向量空间  $\Omega$  中随机选出, 那么在最近邻搜索过程中, 访问数据页面  $mbr_i$  的概率就是  $M\text{Sum}(mbr_i, E[NN^d])$  与向量空间  $\Omega$  相交部分的体积。超球查询转换为点查询的示意图如图 7.1 (a) 所示。

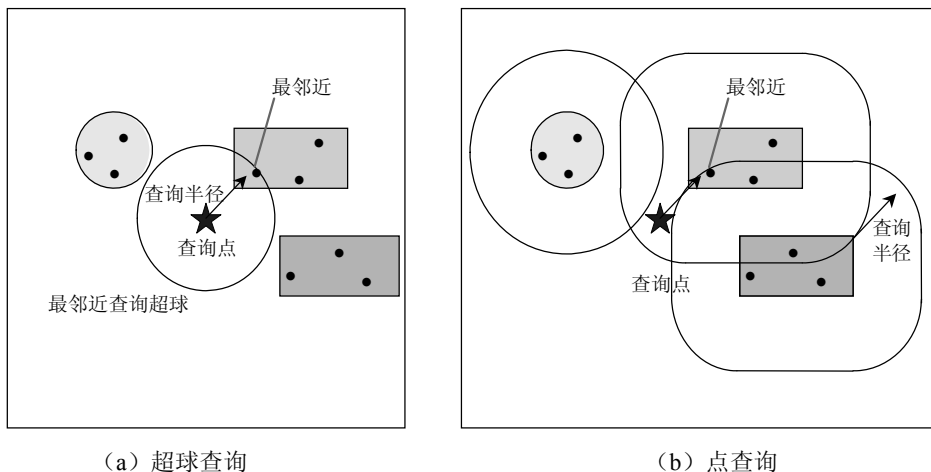


图 7.1 Minkowski 和查询代价模型

### 7.4.2 维数灾难现象的产生

多维索引方法在维数较少的情况下具有很好的检索性能, 但后来的研究和实验却发现, 这些索引方法在维数足够高的情况下, 检索性能下降很快, 其效率甚至会低于最原始的顺序查找方法。依据高维空间中近邻搜索的查询代价模型 (Berchtold S. et al 1997b), Weber 等人 (Weber R. et al 1998) 得出以下结论:

- (1) 任何基于空间分割或数据分割 (聚类) 的索引方法, 存在一个阈值, 当维数大于阈值时, 其近邻搜索性能要低于顺序查找算法。
- (2) 随着维数上升, 基于空间分割或数据分割的索引方法, 其近邻搜索复杂度趋于线性  $O(N)$ 。
- (3) 任何基于空间分割或数据分割 (聚类) 的索引方法, 存在一个阈值, 当维数大于阈值时, 需要访问所有的数据页面。

我们以数据分割的索引方法 R 树为例, 解释维数灾难的产生。R 树的查找算法是一种深度优先算法, 在查询过程中, MBR 区域被用作剪枝搜索过程中的目录结点。以最近邻搜索为例, 如果某个 MBR 与查询超球  $NN^p(q)$  不相交, 那么整个 MBR 中的数据向量都不需要访问, 因为包含在此 MBR 中的所有向量与查询向量之间的距离都大于近邻搜索半径  $NN^d(q)$ 。由于 MBR 可能相互重叠,  $NN^p(q)$  可能与多个 MBR 相交, 无法保证查找操作只搜索一个路径即可成功。

当维数增加时, R 树将发生以下变化: 首先, MBR 的体积将随着维数的增加而增大,

MBR 之间的重叠率将迅速上升, 参考文献 (Berchtold S. 1996) 中指出, 当维数大于 10 时, 对于非均匀分布的数据集, 其 MBR 的重叠率将迅速达到 100%; 其次, 随着维数的增加, 近邻搜索半径也随之增大。近邻搜索距离受维数影响更大, 当维数增加时, 近邻搜索距离迅速增大, 并且很快超过向量空间  $\Omega$  的边界。

由于上述现象的产生, 在进行近邻搜索时, 随着维数的升高, 近邻搜索超球  $NN^q(q)$  将与越来越多的 MBR 相交, 当维数增加到一定程度时, 就需要遍历访问 R 树中所有的路径和结点, 此时其搜索性能肯定会低于顺序查找算法。图 7.2 给出了近邻搜索半径变化的对近邻访问页面的影响 (崔江涛 2005)。

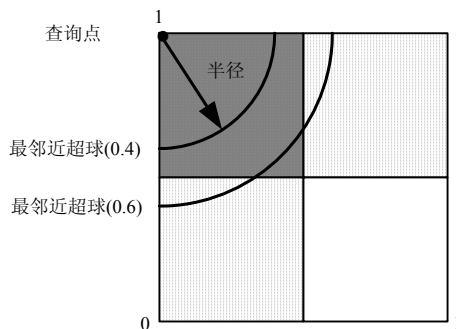


图 7.2 近邻搜索半径对访问区域的影响

另外, 从索引结构的角度来看, 当维数升高时, 表示数据点所需的存储空间也会变大, 数据结点的扇出度会变小, 这会造成建树过程中频繁的分叉以及树的高度增加, 也会严重影响查询效率。

## 7.5 基于矢量量化的向量近似方法

### 7.5.1 矢量量化

矢量量化技术是近年来图像、语音信号编码技术中广泛应用的量化编码方法 (孙圣和等 2002)。它是标量量化的自然发展, 是一种高效的数据压缩技术, 能够有效的提高压缩比。

#### 1. 矢量量化原理

基本的矢量量化器可以定义为从  $n$  维欧氏空间  $R^n$  到有限子集的一个映射, 即  $Q: R^n \rightarrow C$ , 其中  $C = \{y_1, y_2, \dots, y_N | y_i \in R^n\}$  称为码书,  $N$  为码书大小。该映射满足  $Q(x | x \in R^n) = y_i$ , 其中  $x = (x_1, x_2, \dots, x_n)$  为  $R^n$  中的  $n$  维矢量,  $y_i = (y_{i1}, y_{i2}, \dots, y_{in})$  为码书  $C$  中的码字并满足

$$d(x, y_i) = \min_{1 \leq j \leq N} d(x, y_j) \quad (7-4)$$

$d(x, y_i)$  是输入矢量  $x$  与码字  $y_i$  之间的失真测度 (距离)。每一个矢量  $x$  都能在码书中找到其最近码字  $y_i = Q(x | x \in R^n)$ 。输入矢量空间通过量化器  $Q$  量化后, 可以用划分  $S = \{S_1,$

$\mathbf{S}_2, \dots, \mathbf{S}_N$  来描述, 其中  $\mathbf{S}_i$  是所有映射成码字  $\mathbf{y}_i$  的输入矢量集合, 即  $\mathbf{S}_i = \{\mathbf{x} | Q(\mathbf{x}) = \mathbf{y}_i\}$ 。这  $N$  个子空间  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N$  满足

$$\bigcup_{i=1}^N \mathbf{S}_i = \mathbf{S} \text{ 且 } \mathbf{S}_i \cap \mathbf{S}_j = \Phi (i \neq j) \quad (7-5)$$

矢量量化器根据一定的失真测度在码书中搜索出与输入矢量之间失真最小的码字。矢量量化解码过程很简单, 只要根据码字索引在码书中查找该码字, 将其作为重构矢量。

## 2. 码书设计

设计矢量量化器的主要任务是设计码书, 码书设计问题可以用两个优化准则来描述:

### (1) 最近邻条件 (最佳划分)

对于给定码书, 训练矢量集的最佳划分可通过把每个训练矢量映射为离它最近的码字而得到。设训练矢量集为  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ , 则训练矢量集的最佳分类  $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N\}$  满足

$$\mathbf{S}_i = \left\{ \mathbf{v} | d(\mathbf{v}, \mathbf{y}_i) = \min_{1 \leq k \leq N} d(\mathbf{v}, \mathbf{y}_k), \mathbf{v} \in \mathbf{X} \right\} \quad (7-6)$$

通常把这种划分称为 Voronoi 划分 (又称 Dirichlet 划分), 对应的子集  $\mathbf{S}_i$  称为 Voronoi 胞腔, 下面简称胞腔。

### (2) 质心条件 (最佳码书)

对于给定的训练矢量划分, 其对应的最优码书中各码字可以通过求各聚类的质心矢量而得到。对于给定的码书长度  $N$  的情况下, 可以采用 LBG 算法设计最佳量化器 (孙圣和等 2002)。下面给出已知数据点集的矢量量化器设计算法步骤。

#### ① 已知数据集的矢量量化器设计算法

(a) 给定初始码书  $\mathbf{y}_N^{(0)}$ , 即给定码书大小  $N$  和码字  $\{\mathbf{y}_1^{(0)}, \mathbf{y}_2^{(0)}, \dots, \mathbf{y}_N^{(0)}\}$ , 并设置  $k = 0$ , 设起始平均失真  $D^{(-1)} \rightarrow \infty$ , 给定计算停止门限  $\varepsilon (0 < \varepsilon < 1)$ 。

(b) 用码书  $\mathbf{y}_N^{(k)}$  为已知质心, 根据最佳划分原则把数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$  划分为  $N$  个胞腔, 即

$$\mathbf{S}_j^{(k)} = \left\{ \mathbf{v} | d(\mathbf{v}, \mathbf{y}_j^{(k)}) = \min_{1 \leq i \leq N} d(\mathbf{v}, \mathbf{y}_i^{(k)}), \mathbf{v} \in \mathbf{X} \right\} \quad (7-7)$$

(c) 计算平均失真与相对失真。平均失真为

$$D^{(k)} = \frac{1}{m} \sum_{r=1}^m \min_{1 \leq j \leq N} d(\mathbf{x}_r, \mathbf{y}_j^{(k)}) \quad (7-8)$$

判断相对误差是否满足  $\left| (D^{(k-1)} - D^{(k)}) / D^{(k)} \right| \leq \varepsilon$ , 如果满足, 则停止计算, 当前码书  $\mathbf{C}^{(k)}$  就是所求得码书。否则进行步骤 (d)。

(d) 根据最佳码书条件, 计算各胞腔的质心, 即

$$\mathbf{y}_j^{(k+1)} = \frac{1}{|\mathbf{S}_j^{(k)}|} \sum_{\mathbf{v} \in \mathbf{S}_j^{(k)}} \mathbf{v} \quad (7-9)$$

由这  $N$  个新质心  $\{\mathbf{y}_1^{(k+1)}, \mathbf{y}_2^{(k+1)}, \dots, \mathbf{y}_N^{(k+1)}\}$  构成新的码书  $\mathbf{C}^{(k+1)}$ , 并置  $k = k + 1$ , 转步骤 (b)。

LBG 算法中, 初始码书的选择影响码书训练的收敛速度和最终码书的性能。传统的码书选择算法包括随机选择法和分裂法 (孙圣和等 2002)。用分裂法产生的码书性能较好, 下面给出采用分裂法的 LBG 算法。

## ② 分裂 LBG 算法

(a) 计算所有训练矢量的质心  $y_1^{(0)} = \frac{1}{m} \sum_{j=1}^m x_j$ 。

(b) 用一个合适的参数  $A$ ，乘以码字  $y_1^{(0)}$ ，形成第二个码字  $y_2^{(0)}$ 。

(c) 以码字  $y_1^{(0)}$  和  $y_2^{(0)}$  为初始码字，用前面所述的 LBG 算法设计仅含有 2 个码字的码书  $C_2^{(k)} = \{y_1^{(k)}, y_2^{(k)}\}$ 。

(d) 将码书  $C_2^{(k)}$  中的 2 个码字  $y_1^{(k)}$  和  $y_2^{(k)}$  分别乘以合适的参数  $B$ ，得到 4 个码字  $y_1^{(k)}$ ， $y_2^{(k)}$ ， $By_1^{(k)}$ ， $By_2^{(k)}$ 。

(e) 以这 4 个码字为初始码书，用前面所述的 LBG 算法设计仅含有 4 个码字的码书，再对设计好的 4 个码字乘以适当的系数进一步扩大码字的数目。如此反复，经过  $\log_2^N$  次设计，就得到所要求的含  $N$  个码字的初始码书。

除经典的 LBG 算法及其改进算法外，码书生成算法还包括成对最近邻算法（Pairwise Nearest Neighbor, PNN）和最大下降法（Maximum Design, MD），基于神经网络的矢量量化码书设计算法，基于全局优化技术的模拟退火码书设计算法，基于模糊聚类理论的码书设计算法等（陆哲明 2001）。

## 7.5.2 基于矢量量化技术的索引结构（Cui J.T. et al 2007，崔江涛 2005）

设  $p_i (i=1, 2, \dots, n)$  表示特征向量训练集的某一向量，采用分裂 LBG 算法生成矢量量化器  $Q$ 。在进行矢量量化编码时，输入向量  $p_i$  与码书  $C$  中的每一个码字进行比较，分别计算出它们的失真（距离）。搜索到的距离最小的码字  $y_k$  就是  $p_i$  的量化码字， $p_i$  的编码为码字索引  $k$ 。所有向量的近似向量顺序排列，组成近似向量文件，同样，从存储设备读取近似向量时，采用顺序读取的方式。

## 1. 数据组织形式

矢量量化的解码过程中，根据近似向量从码书中查出相应的码字  $y_k$ ， $y_k$  就是  $p_i$  的重构向量，它是  $p_i$  的近似表示。向量近似方法中，可以根据近似向量计算距离上、下界，从而可以对近似向量文件进行过滤。为了能够利用重构向量过滤近似向量文件，需要能够定量表示数据点分布范围，即可以计算出查询向量与原向量的距离上、下界。

在传统的树型索引结构中，目录结点需要描述其子树中所有结点的表示区域，通常采用的描述方式包括 MBR 和超球。由于多个向量可能量化为同一个码字，那么同一个胞腔中可能包含多个数据点，此时可以采用 MBR 或超球来表示同一个超球中多个数据点所表示的区域。

MBR 的构造过程比较简单，一个  $n$  维近似向量生成的 MBR 区域可以表示为

$$I = [g_1, h_1] \times \cdots \times [g_n, h_n]$$

式中， $g_j (1 \leq j \leq n)$  表示包含在该 MBR 区域中的数据点在第  $j$  维空间中的坐标最小值， $h_j (1 \leq j \leq n)$ ，是包含在该 MBR 区域中的数据点在第  $j$  维空间中的坐标最大值。用 MBR 组织胞腔内数据如图 7.3 (a) 所示，每个胞腔对应的 MBR 需要在编码阶段生成超球的构造过程稍微复杂些。由矢量量化过程可知，码字  $y_k$  是其对应胞腔  $S_k$  中所有数据点的质心，设  $r_k$  表示该胞腔对应的超球的半径，则胞腔内的数据点都包含在以  $y_k$  为球心，以  $r_k$  为半径

的超球内。用超球表示胞腔内数据点的示意图如图 7.3 (b) 所示。其中, 矢量量化胞腔半径定义为

$$r_k = \max_{p_i \in S_k} (d(p_i, y_k)) \quad (7-10)$$

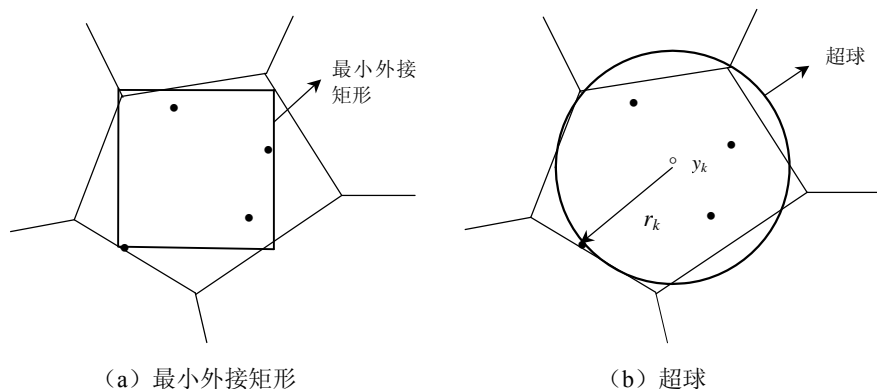


图 7.3 胞腔内数据点区域构造方法

采用 MBR 和超球来作为胞腔区域, 只是一种近似表示。胞腔将整个数据空间划分为互不相连的区域, 采用超球或 MBR 表示后, 不同的胞腔区域之间将会相互重叠, 当胞腔内数据点数量较多时, 交叠现象将会更加明显, 如图 7.4 所示。这在一定程度上降低了近邻搜索算法的过滤效率。

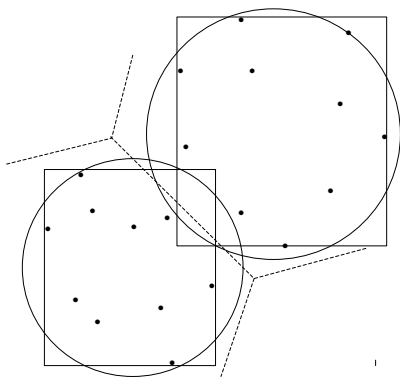


图 7.4 MBR 或超球表示胞腔区域时的重叠现象

## 2. 改进的矢量量化算法

与达到最小失真的目标不同, 矢量量化应用于近邻查询过程, 主要目标是减少原始向量访问量。而只有当每个胞腔内数据点数量大致相同时, 才可以达到较好的查询性能。于是, 在码书设计阶段, 就应该调整设计目标, 使各个胞腔中的数据点数量大致相同。

通过大量的实验可知, 当码书长度增加时, 空胞腔以及仅包含一个数据点的胞腔数量增加很快。例如, 对 10 000 个向量进行矢量量化, 码字索引长度为 8 时, 几乎没有空胞腔存在; 而当码字索引长度为 13 时, 空胞腔数目能够占到胞腔总数目的 1/3。另外, 通过分裂 LBG 矢量量化算法对数据空间进行划分, 胞腔内的数据点数量分布很不均匀, 某些胞腔中会包含有较多数量的数据点。在维数较高的情况下, 采用 MBR 或超球表示这些胞腔区

域时,数据点越多,MBR或超球的体积越大,与查询点的距离下界也会变小,在近邻搜索过程中被访问的概率也相应增大。针对上述现象,我们设计了应用于近邻搜索的改进的分裂LGB算法,改进的分裂LGB算法的基本思想与去除空胞腔分裂法相同。

(1) 对空胞腔进行操作

把空胞腔去掉,将最大的胞腔 $S_m$ 分裂为两个小胞腔。分裂方法如下:

① 用一个合适的参数 $A$ ,乘以码字 $y_m$ ,得到两个码字 $y_{m1} = y_m$ 和 $y_{m2} = Ay_m$ 。

② 以 $y_{m1}$ 和 $y_{m2}$ 这两个码字来划分这个大胞腔,构成两个小胞腔 $S_{m1}$ 和 $S_{m2}$ 。同时可用类似LGB算法的分裂方法更精确的计算 $S_{m1}$ 和 $S_{m2}$ 的质心。

(2) 消除胞腔内数据点较少的胞腔

① 如果胞腔内数据点数量小于一定的阈值,则将该胞腔对应的码字去掉,将数据点重新划分到其他胞腔中。阈值的选择与数据集大小和码书长度都有关系,一般取值低于胞腔中数据点数量的平均值。

② 将最大的胞腔分裂为两个小胞腔,分裂方法与上面相同。

采用上述改进算法,能够使每个胞腔中数据点的分配达到近似最优。实验表明,采用上述改进算法生成的码书能够提高索引结构的近似向量过滤效率。

综上所述,基于矢量量化的近似向量生成算法的过程如下:

(1) 采用改进的LGB方法生成码书;

(2) 根据码书,对数据集中的特征向量 $p_i$ 进行编码,得到其编码码字 $y_k$ 和编码索引 $k$ ;

(3) 根据 $p_i$ 更新胞腔 $S_k$ 的表示区域(MBR或超球);

(4) 所有向量的编码索引进行顺序排列组成向量近似文件;

(5) 存储每个胞腔的表示区域。

### 7.5.3 码书长度分析与乘积码书法

在量化性能方面,矢量量化要明显好于标量量化,但是在编码过程中,矢量量化复杂度要远大于标量量化。矢量量化也受到维数灾难的影响,其复杂度随着维数的增长而呈指数式增加,这也是实现高维矢量量化的主要障碍。另外,随着数据集数目的增大,设计码书的时间也迅速增加。在基于矢量量化的近似近邻搜索过程中,虽然码书的构造和编码过程是离线进行的,但是太高的复杂度仍然会影响其实际应用。在矢量量化过程中,码书长度、训练矢量个数以及维数都会影响量化性能和量化时间。

码书长度的选择是影响量化效果的一个很重要的方面,码书长度过小,量化误差会加大;码书长度增加又会增加码书构造过程和码书搜索过程的复杂度。在编码以及检索过程中,码书要驻留在内存中,但是随着码书长度的加大,内存有可能容纳不下整个码书。在矢量量化过程中,为降低码书存储容量,常用的方法包括乘积码书法和多级矢量量化器。

乘积码书法是一种码书初始化的方法,它是用若干个低维数的码书作为乘积码,求得所需的高维数的码书。比如说,要设计一个高维数的码书,可简单的用2个低维数的码书作为乘积来获得,即 $C_N^{(0)} = C_{N_1}^{(01)} \times C_{N_2}^{(02)}$ ,其中 $C_{N_1}^{(01)}$ 是维数为 $K_1$  ( $K_1 < K$ ),大小为 $N_1$ 的码书, $C_{N_2}^{(02)}$ 是维数为 $K - K_1$ ,大小为 $N_2$ 的码书,这时 $K$ 维码书 $C_N^{(0)}$ 的码书长度为 $N = N_1 \times N_2$ ,但其存储容量为 $N_1 + N_2$ 。图7.5是乘积矢量量化示意图。

采用乘积矢量量化,即可以降低码书生成的复杂度,又可以降低码书存储容量。乘积量化法对量化误差和过滤效率的影响将在实验部分进行详细分析。值得注意的是,乘积量

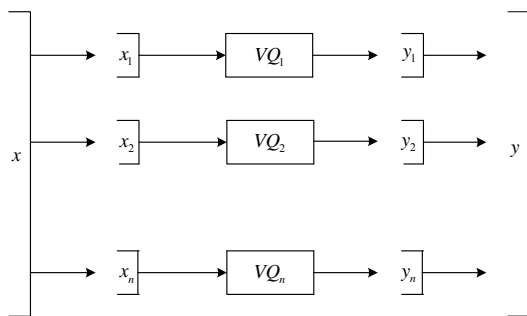


图 7.5 乘积矢量量化示意图

化器的极限情况就是分别对矢量中的每一维分量进行量化，就是标量量化，类似于向量近似方法中采用的方法。

### 7.5.4 近邻搜索算法

矢量量化方法中的近似搜索算法分为两个阶段。第一阶段对近似向量序列进行顺序读取，对特性向量进行过滤；第二阶段对原始向量进行操作，得到最终的搜索结果。对特征向量进行过滤，需要根据近似向量计算特征向量与查询向量之间的距离上、下界。在进行近邻搜索前，首先计算查询向量  $q$  与胞腔  $S_k$  表示区域之间的距离上、下界，并且存储在数组中。因为向量编码的码字索引就是近似向量，所以在访问近似向量时，根据码字索引可以从数组中直接查找到查询向量与胞腔之间的距离上、下界。查询向量与 MBR 或超球之间的距离上、下界的示意图如图 7.6 所示。

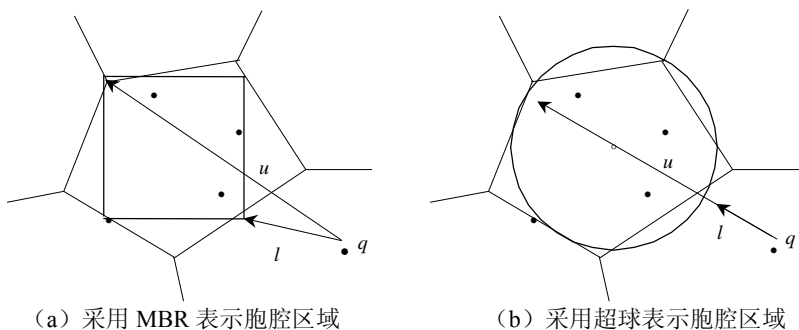


图 7.6 查询向量与 MBR 或超球之间的距离上、下界

胞腔区域用 MBR 表示时，距离下界和上界分别定义为：

$$l_k = \left( \sum_{j=1}^n (l_{k,j})^p \right)^{1/p} \quad (7-11)$$

$$u_k = \left( \sum_{j=1}^n (u_{k,j})^p \right)^{1/p} \quad (7-12)$$

$$\text{式中 } l_{k,j} = \begin{cases} q_j - h_j & q_j > h_j \\ 0 & q_j \in [g_j, h_j] \\ g_j - q_j & q_j < g_j \end{cases}, \quad u_{k,j} = \begin{cases} q_j - h_j & q_j > h_j \\ \max(q_j - g_j, h_j - q_j) & q_j \in [g_j, h_j] \\ g_j - q_j & q_j < g_j \end{cases}。$$

胞腔区域用超球表示时, 距离下界和上界分别定义为:

$$l = \begin{cases} d(\mathbf{q}, \mathbf{y}_k) - r_k & d(\mathbf{q}, \mathbf{y}_k) > r_k \\ 0 & d(\mathbf{q}, \mathbf{y}_k) \leq r_k \end{cases} \quad (7-13)$$

$$u = d(\mathbf{q}, \mathbf{y}_k) + r_k \quad (7-14)$$

**定理 2:** 采用超球表示胞腔区域时, 查询向量与胞腔内数据点的距离介于距离上、下界之间, 即  $l \leq d(\mathbf{q}, \mathbf{p}_i) \leq u$  (证明见崔江涛 2005)。

设数组  $knn\_u$  来保存到目前为止的  $k$  个最近邻向量的标记  $i$  以及它们到查询点  $\mathbf{q}$  的距离上界  $u_i$ , 数组中的数据按升序排列, 则  $k\_NN$  搜索第一阶段算法如下:

- (1) 初始化数组  $knn\_u$ ;
- (2) 计算查询向量与码字之间的距离上、下界, 并将结果存储在数组中;
- (3) 顺序读取向量近似文件, 并且根据近似向量计算与查询向量之间的距离下界  $l_i$ , 如果  $l_i < knn\_u[k]d$ , 则排除该向量; 否则, 计算距离上界  $u_i$ , 如果  $u_i < knn\_u[k]d$ , 则  $knn\_u[k]d$  更新为  $u_i$ , 数组  $knn\_u$  中的元素按照距离重新排序;
- (4) 如果  $i = N$ , 算法结束; 否则  $i = i + 1$ , 继续步骤 3。

### 7.5.5 实验分析

测试数据集是一个航拍图像库, 包含了 275, 245 幅卫星图像, 使用 60 维 Gabor 纹理作为图像特征向量 (Manjunath B.S. et al 1996) (<http://vision.ece.ucsb.edu/datasets/>), 在测试算法特性的实验中, 选择 10000 个特征向量进行测试, 在测试数据集大小对算法的影响时, 再选择 50000 个特征向量进行测试, 测试结果是 100 个随机选择查询向量性能数据的平均值。

#### 1. 胞腔表示区域对过滤性能的影响

实验结果在乘积码书法生成码书的情况下取得, 码书是由 10 个 6 维的码书生成的乘积码书。码书中码字长度分别实现了 6 位, 7 位和 8 位三种情况。图 7.7 (a) 是第一阶段过滤效率, 图 7.7 (b) 是第二阶段过滤效率。

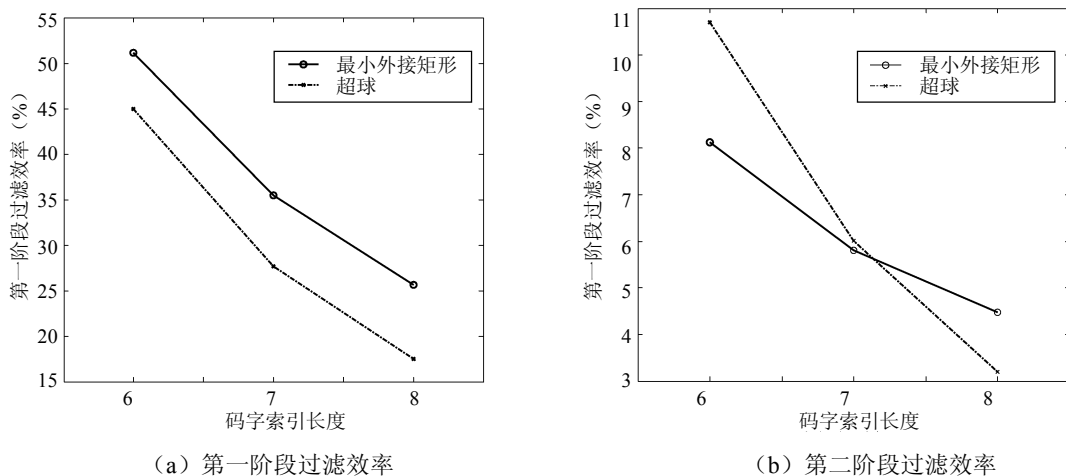


图 7.7 不同超球表示区域的过滤效率比较



由图 7.7 (a) 可以看出, 超球的第一阶段过滤效率要低于 MBR, 在其他生成码书的情况下, 这个结论也是成立的; 由图 7.7 (b) 可以看出, 在较低码字长度的情况下, MBR 的第二阶段过滤效率要低于超球, 在码字长度较高的情况下, 超球的过滤效率要低于 MBR。随着码字长度的升高, 超球的过滤效率下降速度明显高于 MBR, 也就是说, 在码字长度较高的情况下, 超球的第二阶段过滤效率也会低于 MBR。在下面的实验中, 都以超球作为胞腔内数据点的表示区域。

## 2. 码字长度对过滤性能的影响

本节比较了不同码字长度情况下, VQ 方法的过滤效率。图 7.8 (a) 是不采用乘积量化 (常规码书) 的情况, 图 7.8 (b) 是采用乘积量化 ( $6 \times 10$  乘积码书) 的情况, 子码书个数为 10, 每个码书对 6 维向量进行矢量量化。

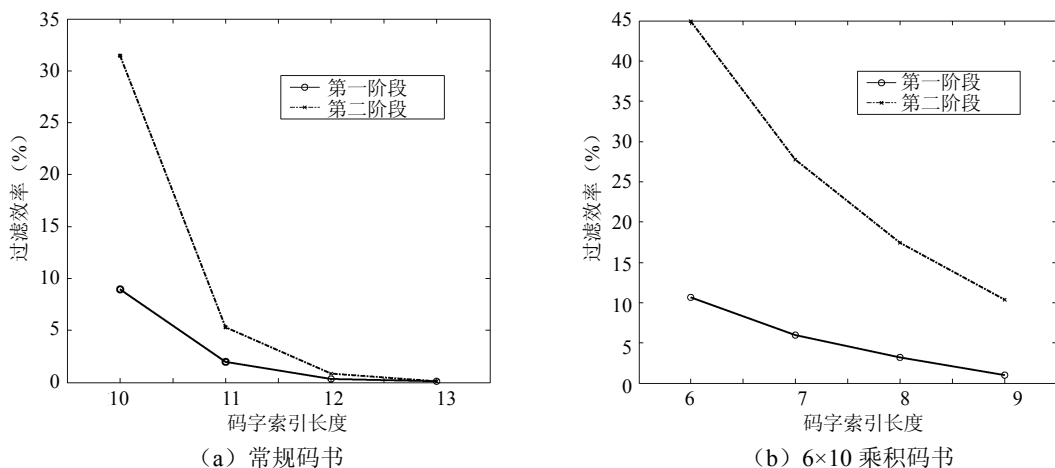


图 7.8 码字索引长度对过滤效率的影响比较

由图 7.8 可以看出, 码字长度越长, 两个阶段的过滤能力越强。码字长度为 13 时, 大部分胞腔中只有一个数据点, 其累计过滤效率达到最高, 但其码书容量也与原始向量文件相差无几, 所以并无实际应用价值。在不采用乘积码书的情况下, 随着码字长度的降低, 过滤效率下降很快, 这主要是由每个胞腔中的平均数据量增大引起的。胞腔中数据点增多, 不但更多的向量会被量化为同一个码字, 而且也会增加胞腔表示区域超球的重叠现象, 查询点往往包含在多个超球之中, 这种重叠现象会严重影响 VQ 方法的过滤效率。

## 3. 乘积码书数量对过滤性能的影响

本节测试了在乘积矢量量化中码书数量对过滤效率的影响。对于固定的码字长度, 乘积码书的数量对整个码书的存储容量并无影响, 但是码书数量越多, 矢量量化过程中计算的维数越低, 矢量量化的复杂度越低。在极限情况下, 就是标量化情况下, 其复杂度达到最低。另外, 近似向量的表示长度与码书数量有关, 在码字索引长度不变的情况下, 近似向量文件的长度与码书数量呈正比关系。图 7.9 给出了码字索引长度为 8 位的情况下, 乘积码书数量对 VQ 方法过滤效率的影响。

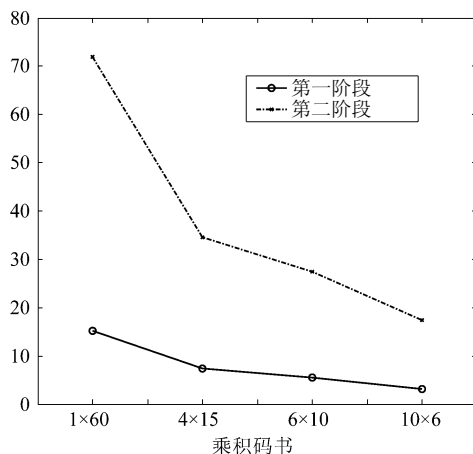


图 7.9 乘积码书数量对 VQ 方法过滤效率的影响

对上述结果进行分析，乘积码书数量越多，过滤效率就越高，也就是说在标量量化情况下，过滤效率达到最佳。可以从胞腔数目来解释这一现象，假定码字索引长度为  $b$ ，那么不采用乘积量化情况下，矢量量化胞腔的数目为  $2^b$ ，如果采用乘积矢量量化，码书数目为  $m$ ，那么整个数据空间内被分割为  $2^{b*m}$  个胞腔。在对 VA 方法的分析中可知，数据空间被分割的越精细，过滤性能越好在 VQ 方法中此原则仍然适用。所以乘积码书数目越多，分割数据空间的胞腔数目越多，过滤效率也就越高。

#### 4. 数据集大小对过滤性能的影响

本节测试了数据集大小对 VQ 方法过滤性能的影响。图 7.10 给出了两种矢量量化情况下 VQ 方法的第二阶段过滤性能。可以看出，在不采用乘积码书的情况下，VQ 方法的过滤效率受数据集大小影响较大，数据集越大过滤效率越低。随着数据集的增大，胞腔内的数据点数目也相应增多，用于表示胞腔内数据点的超球体积也相应增大，而与查询向量之间的距离下界将会减小，这都会导致过滤效率的迅速下降。在乘积码书情况下，VQ 方法的过滤效率受数据集大小影响较小。采用较高的码字索引长度时，随着数据集的增大，过滤效率会逐渐降低，也就是说，原始向量的访问数量并不随着数据集的增长而线性增长。

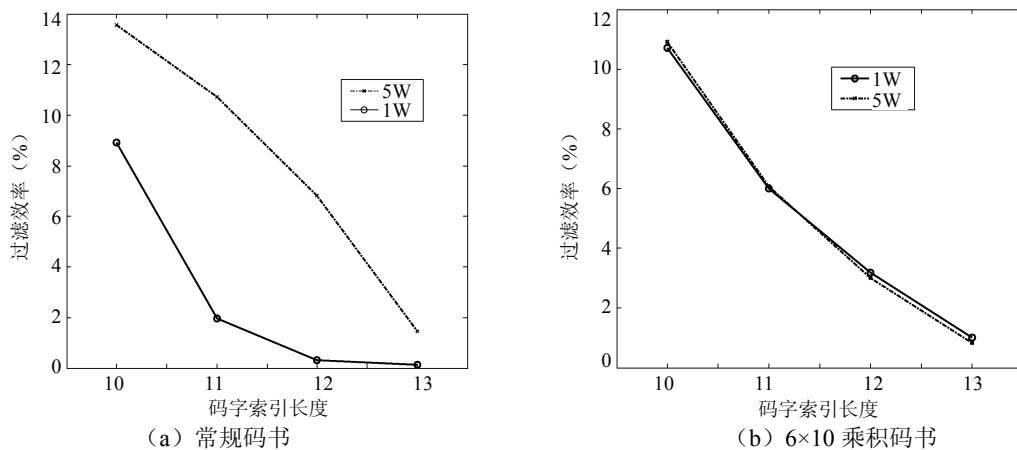


图 7.10 数据集大小对第二阶段过滤效率的影响

### 5. VA方法与VQ方法性能比较与分析

在相同数据集上对 VA 方法和 VQ 方法进行比较。首先比较 CPU 运算代价,只考虑两种方法对近似向量进行运算的 CPU 时间。VA 和 VQ 方法中计算查询向量与特征向量的距离上、下界时需要进行查找数组运算,VA 方法在所有分量上进行数组查找并求和,而 VQ 方法只需对几个数组进行操作即可(乘积量化中,数组数目与子向量数目相同)。所以,VQ 方法的 CPU 运算时间要略低于 VA 方法,实验中也证明了这一点。

I/O 复杂度是评价索引结构的主要因素,由上面实验中对 VQ 方法的分析可知,采用乘积码书法,采用合适的码字索引长度,可以达到较好的过滤能力。例如,采用  $10 \times 6$  乘积码书,码字索引长度为 9 时的累计过滤能力与 VA 方法中近似位数为 4 时的过滤能力相当,但是 VQ 方法中近似文件大小只是 VA 方法的一半大小。如果不计码书的存储和 I/O 代价,选择合适的码书数量和码字索引长度,VQ 方法的性能总能高于 VA 方法,这种性能的提升主要靠降低向量近似文件来获得的,数据集越大,VQ 方法性能提升越大。

进一步对 VA 方法进行分析,当每维近似位数为 4 时,数据空间被划分为  $2^{60 \times 4}$  个胞腔,胞腔数目已经远远大于数据集大小,数据点在胞腔中是及其稀疏分布的,大部分的胞腔都是空胞腔,近似向量的表示存在一定的冗余。而矢量量化方法则较好的解决了这个问题,可以用较少的胞腔来划分整个数据空间。矢量量化的一个较难克服的缺点就是胞腔形状难以定量描述,采用超球近似描述胞腔区域时,相邻的超球会产生重叠现象,这在一定程度上减弱了其过滤能力,但是这种现象在 VA 方法中并不存在。

## 7.6 二次式距离上基于SVD的高维索引方法

在 CBIR 系统中,除欧氏距离外,常用的距离度量方式还包括二次式距离。由于二次式距离的定义中考虑了分量之间的相关性,许多系统中都采用了二次式距离度量方式。但是 VA 方法并不直接支持二次式距离,本节在引入奇异值分解(Singular Value Decomposition, SVD)技术的基础上,讨论基于二次式距离的 VA 方法(崔江涛等 2006b)。

### 7.6.1 奇异值分解

向量  $p$  和  $q$  之间的二次式距离可以表示为

$$d(q, p) = (q - p)^T A (q - p) \quad (7-15)$$

式中,  $A = (a_{i,j})_{i,j=1,\dots,n}$  是相似矩阵。在 CBIR 系统中  $A$  通常是一个正定对称矩阵。 $a_{i,j}$  是分量  $i$  和  $j$  之间的相似系数。我们假设,  $a_{i,i} = 1$ ,  $0 \leq a_{i,j} < 1$  ( $i \neq j$ )。

根据矩阵的 SVD 技术,对于正定对称矩阵  $A$ ,存在正交矩阵  $V$ ,使得  $A = V^T \Sigma V$ 。其中  $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d)$ ,并且  $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_n \geq 0$ ,  $\lambda_i$  是矩阵  $A$  的奇异值,也是  $A$  的特征值。经过 SVD 后,二次式距离可以表示为

$$d(q, p) = (Vq - Vp)^T \Sigma (Vq - Vp) \quad (7-16)$$

这样二次式距离可以表示成加权欧氏距离形式,而  $\lambda_i$  就是各分量的能量加权系数,上、下界距离计算公式为:

$$l_i = \left( \sum_{j=1}^n \lambda_j (l_{i,j})^2 \right)^{1/2} \quad (7-17)$$

$$u_i = \left( \sum_{j=1}^n \lambda_j (u_{i,j})^2 \right)^{1/2} \quad (7-18)$$

假设变换后第  $i$  维分量的方差为  $\sigma_i^2$ ，将加权系数考虑在内，式 (7-17) 可以表示为

$$d(\mathbf{q}, \mathbf{p}) = (\mathbf{A}\mathbf{V}\mathbf{p} - \mathbf{A}\mathbf{V}\mathbf{q})^T (\mathbf{A}\mathbf{V}\mathbf{p} - \mathbf{A}\mathbf{V}\mathbf{q}) \quad (7-19)$$

式中， $\mathbf{A} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$ ，那么第  $i$  维分量的方差为  $\lambda_i \sigma_i^2$ 。在建立近似向量的过程中，依然采用非均匀的近似位数分配算法和等密度分割算法。

## 7.6.2 索引结构

根据主分量排序思想，在二次式距离上也根据能量最大的分量来进行排序，建立排序后的向量近似文件。经过 SVD 变换后，向量各分量之间并不具备多分辨率特性，在  $k$  近邻搜索过程中，采用了类似的低维过滤算法，即首先在低维空间上计算距离，判断是否已经超过近邻搜索半径，如果超过，则排除此向量，避免进行高维运算。但新的低维过滤算法与 QBIC 系统中的低维过滤算法有所区别。对于一个  $n$  维向量，假定选定的低维维数为  $s$ ， $1 \leq s \leq n$ 。新算法中选择方差最大的  $s$  个分量计算距离来进行过滤。文献 (Hafner J. et al 1995) 中指出，在未知数据分布的情况下，采用加权系数最大  $s$  个分量来计算距离，可以达到对原距离最好的近似。在已知数据集的情况下，选择方差最大的  $s$  个低维分量进行过滤，其过滤效果是最优的。

二次式距离上近似向量文件建立算法步骤如下：

- (1) 对相似矩阵  $\mathbf{A}$  进行 SVD 分解，计算加权系数矩阵  $\mathbf{\Sigma}$  和变换矩阵  $\mathbf{V}$ ，对原特征向量进行变换；
- (2) 根据各分量方差确定各维空间的近似位串长度；
- (3) 根据方差最大值对变换后的向量进行排序；
- (4) 采用等密度分割算法，对向量进行近似得到向量近似文件；
- (5) 计算每个数据页面中的主分量值的范围，用  $B^+$  树管理这些数据。

## 7.6.3 近邻搜索算法

采用低维过滤算法的第一阶段  $k$  近邻查询过程如下：

- (1) 初始化距离上界数组  $knn\_u$ ；
- (2) 对查询向量  $\mathbf{q}$  进行 SVD 变换，根据  $B^+$  树查找初始访问数据页面；
- (3) 在主分量上计算  $\mathbf{p}_i$  与查询向量  $\mathbf{q}$  之间的距离下界  $l_i^l$ ，如果  $l_i^l > knn\_u[k]d$ ，算法结束。否则在方差最大的  $s$  维分量上计算  $\mathbf{p}_i$  与查询向量  $\mathbf{q}$  之间的距离下界  $l_i^s$ ，若  $l_i^s > knn\_u[k]d$ ，排除向量  $\mathbf{p}_i$ ，否则在所有分量上计算  $l_i$ ，若  $l_i \leq knn\_u[k]d$ ，则计算  $u_i$ ，如果  $u_i \leq knn\_u[k]d$ ，则  $knn\_u[k]d$  更新为  $u_i$ ；

- (4) 如果  $i$  等于  $N$ ，则升序访问结束；否则， $i$  值加 1；执行步骤 3。

$k$  近邻搜索过程的第二阶段过滤过程与 VA 方法相同。

## 7.7 多分辨率高维索引方法

### 7.7.1 基本原理

离散小波变换是一种有效的多分辨率正交变换方法。假设一个  $2^L$  维向量  $\mathbf{X}$ ，在分辨率  $l+1$  ( $l=L-1, \dots, 0$ ) 上的向量  $\mathbf{X}^{l+1}$  可以通过小波变换分解为分辨率  $l$  上的一个低频分量  $\mathbf{X}_A^l$  和低频分量  $\mathbf{X}_D^l$ 。用  $X_A^l(i)$  表示  $\mathbf{X}_A^l$  中的元素，用  $X_D^l(i)$  表示  $\mathbf{X}_D^l$  中的元素，可以得出 (Chaur H.H et al 2000)

$$X_A^l(i) = \langle \mathbf{X}_A^{l+1}, 2^{-(l/2)} \phi((t/2^l) - i) \rangle \quad i=0, \dots, 2^l - 1 \quad (7-20)$$

$$X_D^l(i) = \langle \mathbf{X}_D^{l+1}, 2^{-(l/2)} \varphi((t/2^l) - i) \rangle \quad i=0, \dots, 2^l - 1 \quad (7-21)$$

小波分析中有多种类型的小波变换，从运算复杂度上考虑，我们采用最简单的哈尔小波变换。哈尔小波变换的示意图如图 7.11 所示。哈尔小波的尺度函数和小波函数如下所示

$$\phi(t) = \begin{cases} 1 & t=0,1 \\ 0 & \text{其他} \end{cases} \quad (7-22)$$

$$\varphi(t) = \begin{cases} 1 & t=0 \\ -1 & t=1 \\ 0 & \text{其他} \end{cases} \quad (7-23)$$

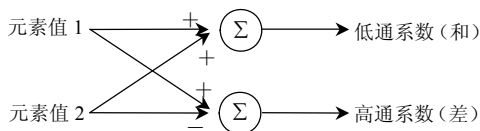


图 7.11 离散哈尔小波变换示意图

一个 8 维向量  $\mathbf{p}$  的分解过程和小波变换域向量如图 7.12 所示。

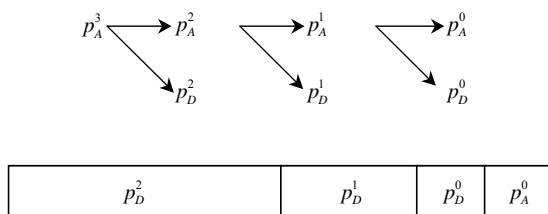


图 7.12 8 维向量  $\mathbf{p}$  的哈尔小波分解过程

根据离散小波变换特性，可以得

$$L_2(\mathbf{Q}_A^l, \mathbf{P}_A^l)^2 = L_2(\mathbf{Q}_A^{l-1}, \mathbf{P}_A^{l-1})^2 + L_2(\mathbf{Q}_D^{l-1}, \mathbf{P}_D^{l-1})^2 \quad (7-24)$$

进一步，可以得到下述不等式：

$$L_2(\mathbf{Q}^L, \mathbf{P}_i^L) \geq \dots \geq L_2(\mathbf{Q}_A^l, \mathbf{P}_A^l) \geq \dots \geq L_2(\mathbf{Q}_A^0, \mathbf{P}_A^0) \quad (7-25)$$

### 7.7.2 索引结构

由于特征向量的相邻分量之间往往存在很大的冗余性, 所以小波变换可以将能量集中在少数几个低频分量上。将 DWT 变换的第一维分量作为主分量, 首先在主分量上计算查询向量与近似向量的距离下界, 如果主分量上计算所得距离下界大于目前第  $k$  小的距离上界, 那么可以直接排除此向量; 进一步, 我们按照特征向量的主分量值对近似向量进行排序, 如果当前访问的向量可以根据主分量上的距离下界排除掉, 那么主分量值更大或更小的近似向量也可以被排除掉, 这样就可以降低近似向量的访问数量。

对数据集进行离散小波变换后, 按照主分量的值对原始向量进行排序, 然后生成近似向量, 在生成近似向量的过程中, 需要进行非均匀近似位数分配和等密度分割算法。将近似向量顺序存储在多个数据页面中, 每个数据页面中主分量值的范围可以用一个二维数组表示, 如  $[a, b]$  表示主分量值范围在  $a$  和  $b$  之间。由于这些二维数组不存在交叉情况, 可以采用  $B^+$  树来管理这些数据。基于主分量排序的索引结构的一个示意图如图 7.13 所示。

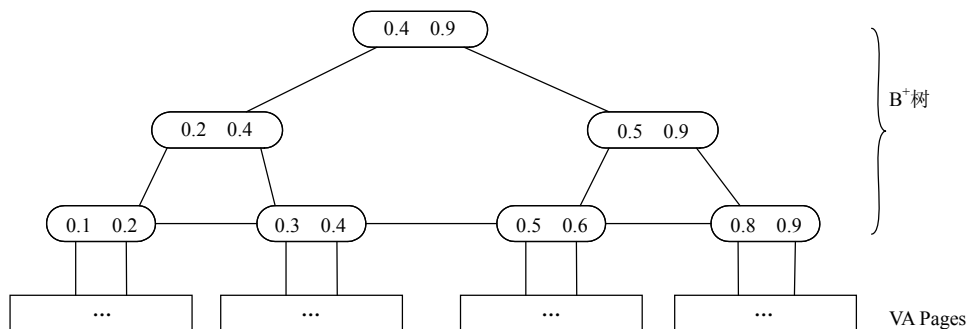


图 7.13 基于主分量排序的索引结构

顺序存储的索引结构存在一个缺点, 就是当数据库中的数据进行更新, 比如插入或删除操作时, 索引结构需要重新建立。一种解决途径就是设定数据页面中数据利用率小于 100%, 即每个数据页面中预留一些空间进行数据更新。当删除特征向量时, 直接从相应的数据页面中删除数据, 并且更新  $B^+$  树即可。当插入新的特征向量时, 可以在数据页面的预留空间中加入新的近似向量, 不需要改动太多的数据页面就可以完成更新过程。当数据更新量较大时, 再对整个索引结构进行改动。主分量排序的索引结构建立过程如下:

- (1) 对原始特征向量进行离散小波变换, 并且按照第一维系数对原始向量进行排序并进行顺序存储;
- (2) 对原始向量序列进行近似, 生成近似向量文件;
- (3) 计算每个数据页面中的主分量值的范围, 用  $B^+$  树管理这些数据。

### 7.7.3 近邻搜索算法

离散小波变换具有多分辨率特性, 分辨率越低, 单位系数上集中的能量越高, 在这些系数上可以利用部分失真搜索 (Partial Distortion Search, PDS) 算法 (叶航军 2003) 高效的排除近似向量。PDS 算法主要用于矢量量化中的码字搜索算法, 这里借鉴其思想, 将其

应用于高维索引的近邻搜索算法中。

PDS 算法的基本思想是：在计算某个近似向量与查询向量的距离下界的过程中，始终判断累积的距离下界是否已经超过目前的近邻距离上界，一旦超出则终止该近似向量与查询向量之间的距离下界计算。 $l_i = \sum_{j=1}^d l_{i,j}$ ，令  $l_i^m = \sum_{j=1}^m l_{i,j}$  为查询向量  $Q$  与近似向量间的部分距离下界，小波变换域的 PDS 算法基本原理如下。

**定理 3：**假定目前第  $k$  小近邻距离为  $knn\_u[k]d$

若  $l_i^m > knn\_u[k]d$  ( $1 \leq m \leq n$ )

则  $l_i \geq l_i^m > knn\_u[k]d$  成立。

根据上述定理，可以在不进行高维计算的情况下，将向量  $p_i$  排除掉，从而降低算法的运算复杂度。初始  $m$  值为 1，如果  $l_i^m > knn\_u[k]d$ ，则排除向量  $p_i$ ；否则采用如下 PDS 算法，从  $m=2$  开始，对每一个  $m$ ， $m=2, \dots, 2^L$ ，计算  $l_i^m$ 。若  $l_i^m > knn\_u[k]d$ ，则排除该向量；否则继续选择下一个  $m$  值进行判断。这个过程直到向量被排除或  $m=2^L$  为止。当  $m=2^L$  时，若  $l_i^m \leq knn\_u[k]d$ ，则计算  $u_i$ ，如果  $u_i < knn\_u[k]d$ ，则  $knn\_u[k]d$  更新为  $u_i$ 。

在矢量量化中的变换域 PDS 算法中，从初始匹配码字开始，采用升序和降序交叉访问的次序，由于码书存储在内存中，这种访问方式是最佳访问方式。如果从外存中读取近似向量，由文献 (Berchtold S. et al 2000) 提供的外存访问模型可知，如果交叉数据页面，将会额外增加页面搜索时间。所以在访问数据页面时，首先在升序方向上顺序访问数据页面，当升序访问结束后，再从初始访问页面开始，降序顺序访问数据页面。这样，新的索引结构就保持了向量近似方法的顺序访问特性。

与其他分量相比，主分量上占据了较大部分能量，在主分量上过滤数据可以有效地限定近似文件的访问范围。下面以升序访问为例，给出基于主分量排序的 VA 方法的  $k$  近邻搜索算法。 $k$  近邻搜索第一阶段过滤算法（升序访问）过程如下：

- (1) 初始化距离上界数组  $knn\_u[k]$ ；
- (2) 计算查询向量  $q$  的变换域向量  $Q$ ，查找初始访问页面；
- (3) 根据目前访问的近似向量  $a_i$  计算  $l_i^1$ 。如果  $l_i^1 > knn\_u[k]d$ ，算法结束。否则，计算  $l_i^m$ ，从  $m=2$  开始，对每一个  $m$ ， $m=2, \dots, 2^L$ ，计算  $l_i^m$ 。若  $l_i^m > knn\_u[k]d$ ，则排除该向量；否则继续选择下一个  $m$  值进行判断。这个过程直到向量被排除或  $m=2^L$  为止。当  $m=2^L$  时，若  $l_i^m \leq knn\_u[k]d$ ，则计算  $u_i$ ，如果  $u_i \leq knn\_u[k]d$ ，则  $knn\_u[k]d$  更新为  $u_i$ ，向量  $p_i$  成为候选向量，进入步骤 4；

- (4) 如果  $i$  等于  $N$ ，则升序访问结束；否则， $i$  值加 1；执行步骤 3；

第二阶段过滤过程与 VA<sup>+</sup>方法 (Ferhatosmanoglu H. et al 2000) 相同，需要访问 DWT 变换域原始特征向量。对第一阶段过滤后的候选向量按照距离下界做升序访问，并分别计算出到查询向量的精确距离。并不是所有的候选向量都需要访问，如果某个候选向量的距离下界超过了目前的第  $k$  个最近邻距离，整个查询过程结束。

## 7.8 向量近似方法在相关反馈技术中的应用

在 CBIR 系统中，常用的相关反馈技术包括两类，一类是修改查询向量或距离判别标

准的方法 (Rui Y. et al 1998, Rui Y. et al 2000), 另一类是将机器学习理论用于相关反馈过程中 (Tong S. et al 2001, Zhang C. et al 2002, Wang J.Z. 2002, Chen Y. et al 2001, Heisterkamp D. et al 2001)。从相似度量度的角度来看, 前者主要采用二次式距离方法, 而后者常采用核函数方法。

### 7.8.1 二次式距离方法

在 CBIR 系统中, 图像间的相似性查询可以看成是向量空间中的  $k$  近邻搜索问题。从向量模型的角度出发, 常用的相关反馈技术包括查询向量优化算法和特征权重调整算法。在相关反馈图像检索系统 MARS (Rui Y. et al 1998) 和 MindReader (Tong S. et al 2001) 中都采用了这两种反馈技术。

假定图像  $i$  的特征向量可以表示为  $\mathbf{p}_i = [p_{i1}, p_{i2}, \dots, p_{in}]$ , 查询特征向量为  $\mathbf{q} = [q_1, q_2, \dots, q_n]$ ,  $n$  是特征向量的维数。  $\mathbf{p}_i$  与  $\mathbf{q}$  之间的距离定义为

$$d(\mathbf{q}, \mathbf{p}_i) = (\mathbf{q} - \mathbf{p}_i)^T \mathbf{W} (\mathbf{q} - \mathbf{p}_i) \quad (7-26)$$

$\mathbf{W}$  称为相似矩阵, 它是一个实对称矩阵, 通过奇异值分解技术,  $\mathbf{W}$  可以变换为一个对角阵, 上述距离就变换为加权欧氏距离形式, 这里假定  $\mathbf{W}$  是一个实对角阵, 并且  $\sum_{jj} \mathbf{W}_{jj} = 1$ 。

假设用户对训练图像标记为正的图像数目为  $N$ , 为使查询向量和反馈正例的距离达到最小, 经过反馈过程后, 查询向量  $\mathbf{Q}$  和相似矩阵  $\mathbf{W}$  的最优表达式为 (Rui Y. et al 1998):

$$\mathbf{Q}^{T*} = \frac{\bar{\pi}^T \mathbf{X}}{\sum_{n=1}^N \pi_n} \quad (7-27)$$

$$\mathbf{W}^* = (\det(\mathbf{C}))^{\frac{1}{d}} \mathbf{C}^{-1} \quad (7-28)$$

式中,  $\pi_n$  是用户对反馈正例图像  $n$  给出的反馈分值。矩阵  $\mathbf{X}$  是由反馈正例向量组成的  $N \times d$  矩阵,  $\mathbf{C}$  是  $\mathbf{X}$  的加权协方差矩阵, 即

$$\mathbf{C}_{r,s} = \frac{\sum_{n=1}^N \pi_n (x_{nr} - q_r)(x_{ns} - q_s)}{\sum_{n=1}^N \pi_n} \quad r, s = 1, \dots, d \quad (7-29)$$

本章 7.5 节提出了基于奇异值分解的二次式距离索引方法, 在搜索过程中采用低维过滤算法, 可以大幅度提高近邻搜索效率。但是这种基于奇异值分解的二次式距离索引方法并不适用于相关反馈过程, 在反馈过程后, 相似矩阵会发生变化, 原有的在变换域建立的近似向量将不再适用, 需要重新建立索引结构才能支持新一轮次的查询过程, 在实际应用中, 这是不现实的。下面给出另外一种支持二次式距离应用的向量近似方法, 该方法能够支持基于特征权重调整的相关反馈过程。

对特征向量进行近似时, 不对向量进行任何变换, 直接生成近似向量。因为每维分量方差相差不大, 所以每维分量分配相同的近似位数。对坐标轴进行分割时仍然采用等密度分割算法。特征向量  $\mathbf{p}_i$  对应的近似向量为  $\mathbf{a}_i$ , 根据式 (7-11) 和式 (7-12) 计算查询向量  $\mathbf{q}$  在每一维分量上的距离上界  $u_{i,j}$  和距离下界  $l_{i,j}$ 。采用二次式距离度量方法, 查询向量  $\mathbf{q}$  与  $\mathbf{p}_i$  的距离上、下界为:

$$\mathbf{l}_i = [l_{i,1}, l_{i,2}, \dots, l_{i,d}] \mathbf{W} [l_{i,1}, l_{i,2}, \dots, l_{i,d}]^T \quad (7-30)$$

$$\mathbf{u}_i = [u_{i,1}, u_{i,2}, \dots, u_{i,d}] \mathbf{W} [u_{i,1}, u_{i,2}, \dots, u_{i,d}]^T \quad (7-31)$$



很容易证明 (崔江涛 2005)

$$l_i \leq d(q, p_i) \leq u_i$$

## 7.8.2 核函数方法

在 CBIR 的相关反馈技术中, 已经证明核函数方法的性能要优于二次式距离方法 (Chen Y. et al 2001, Heisterkamp D. et al 2001)。核函数方法主要解决输入空间非线性可分的问题。使用一个非线性函数  $\phi$ , 将数据映射到一个高维特征空间, 假定  $\phi(x)$  是数据点  $x$  从输入空间到高维特征空间的一个映射。支持向量机理论不直接使用函数  $\phi$ , 而使用高维特征空间的点积运算  $k(a, b) = [\phi(a), \phi(b)]$  来求取距离。 $k$  称为核函数。常用的核函数包括高斯核函数式 (7-32) 和多项式核函数式 (7-33), 即

$$k(a, b) = e^{-\frac{\|a-b\|^2}{2\sigma^2}} \quad (7-32)$$

$$k(a, b) = (1 + \langle a, b \rangle)^d \quad (7-33)$$

高维特征空间中的距离可以通过核函数方式来定义, 对于输入空间中的数据点  $a$  和  $b$ , 有

$$d(a, b)^2 = \|\phi(a) - \phi(b)\|^2 = k(a, a) - 2k(a, b) + k(b, b) \quad (7-34)$$

在 CBIR 系统的相关反馈应用中, 常用的核距离是 AQK (Adaptive Quasiconformal Kernel) 和一类 SVM (One-class Support Vector Machine) (Chen Y. et al 2001, Heisterkamp D. et al 2001)。

AQK 距离中的核函数定义为

$$\tilde{k}(a, b) = c(a)c(b)k(a, b) \quad (7-35)$$

AQK 核距离定义为

$$\begin{aligned} d(q, p)^2 &= \tilde{k}(q, q) - 2\tilde{k}(q, p) + \tilde{k}(p, p) \\ &= c(q)^2 k(q, q) - 2c(q)c(p)k(q, p) + c(p)^2 k(p, p) \end{aligned} \quad (7-36)$$

这里  $c(p)$  是原始输入空间中向量  $p$  的一个正值函数。在反馈过程中, 对于不相关样本可以增大  $c(p)$  来放大空间分辨率, 对于相关样本, 则降低空间分辨率。这样, 不相关图像与查询图像的距离将增大, 相关图像与查询之间的距离将减少。

一类 SVM 就是在训练过程中采用的样本都是正相关样本。经过反馈过程后, 采用一个最小超球来包含大多数的训练样本。一类 SVM 核距离就是样本与超球球心之间的距离。超球球心  $c = \sum_i \gamma_i \phi(x_i)$ , 向量  $z$  与球心之间的核距离为

$$d(z, c)^2 = k(z, z) - 2\sum_i \gamma_i k(x_i, z) + \sum_{i,j} \gamma_i \gamma_j k(x_i, x_j) \quad (7-37)$$

采用 Lagrange 乘子法引入乘子  $\gamma_i$ , 求解下述二次规划问题, 即

$$\min_{R \in \mathbb{R}, \xi \in \mathbb{R}^l} R^2 + \frac{1}{vl} \sum_i \xi_i \quad s.t. \quad \|\phi(x_i) - c\|^2 \leq R^2 + \xi, \quad \xi_i \geq 0 \quad (7-38)$$

那么超球半径与其包含样本数目可由参数  $v \in [0, 1]$  来调节。

文献 (Peng J. et al 2002) 首先采用 M 树来对高维特征空间进行索引, 但是, 在相关反馈过程中, 如果核距离定义发生改变, 需要重新建立 M 树来支持新一轮次的查询过程, 所

以  $M$  树不能实际应用于相关反馈过程。Heisterkamp 等人对核函数方法和向量近似方法进行了深入研究,提出了支持核距离的向量近似方法—KVA-file (Kernel Vector Approxiamtion file) 方法 (Heisterkamp D.R. et al 2003)。

KVA-file 方法在高维特征空间中采用子空间方法建立向量近似文件。首先在特征空间中找到一组压缩的正交向量基,将数据点投影到子空间上建立近似文件,另外,误差(正交补)也作为近似的一部分。假定子空间内正交基集合为  $v_i$ , 其正交补为  $\phi^\perp(x)$ , 这样一个特征空间中的点可以表示为正交基向量与正交补的线性组合。数据点  $P$  和  $Q$  可分别表示为:

$$P = \phi(x_p) = \phi^\perp(x_p) + \sum_{i=1}^n \alpha_i v_i \quad (7-39)$$

$$Q = \phi(x_q) = \phi^\perp(x_q) + \sum_{i=1}^n \beta_i v_i \quad (7-40)$$

式中,  $n$  是子空间的维数(子空间内正交基数目)。 $P$  和  $Q$  之间的核距离可以表示为

$$d(Q, P)^2 = \phi^\perp(x_q)^T \phi^\perp(x_q) + \phi^\perp(x_p)^T \phi^\perp(x_p) - 2\phi^\perp(x_q)^T \phi^\perp(x_p) + \sum_{i=1}^d (\alpha_i - \beta_i)^2 \quad (7-41)$$

对子空间内正交基上的系数  $\alpha$ 、误差  $\phi^\perp(x)^T \phi^\perp(x)$  采用向量近似方法进行近似, 分别建立近似向量文件。式(7-41)中未知的项是  $\phi^\perp(x_q)^T \phi^\perp(x_p)$ 。该项可以表示为:

$$\sqrt{\phi^\perp(x_q)^T \phi^\perp(x_q)} \sqrt{\phi^\perp(x_p)^T \phi^\perp(x_p)} \cos \theta \quad (7-42)$$

$\theta$  是两个向量之间的夹角。夹角信息没有包含在近似向量中, 不过可以根据夹角的极限值求出上述距离的上界和下界。令  $\hat{G}_n(p, p) = \phi^\perp(x_p)^T \phi^\perp(x_p)$ ,  $\hat{G}_n(q, q) = \phi^\perp(x_q)^T \phi^\perp(x_q)$ , 那么距离上界  $u(Q, P)$  和距离下界  $l(Q, P)$  为:

$$u(Q, P)^2 = \hat{G}_n(q, q) + \hat{G}_n(p, p) + 2\sqrt{\hat{G}_n(q, q)}\sqrt{\hat{G}_n(p, p)} + \sum_{i=1}^n (\alpha_i - \beta_i)^2 \quad (7-43)$$

$$l(Q, P)^2 = \hat{G}_n(q, q) + \hat{G}_n(p, p) - 2\sqrt{\hat{G}_n(q, q)}\sqrt{\hat{G}_n(p, p)} + \sum_{i=1}^n (\alpha_i - \beta_i)^2 \quad (7-44)$$

每个  $\alpha_i$  值和  $\hat{G}_n(p, p)$  值可以根据向量近似方法表示成近似向量。该方法实现了两种层次上的压缩近似, 首先是子空间维数, 其次是每维空间分配的近似位数。在该方法中采用非均匀位串分配算法和等密度坐标轴分割算法可以取得更好的性能提升。

### 7.8.3 改进的近邻搜索算法

向量近似方法的近邻搜索包括两个过程: 第一阶段对近似向量进行过滤, 没有被排除掉的向量构成候选向量集, 用  $N_1$  表示; 第二阶段中对原始向量进行操作, 需要访问的原始特征向量的集合用  $N_2$  表示。

在过滤阶段过滤掉更多的向量, 即减少  $N_1$  和  $N_2$  中特征向量的数目, 就可以获得更好的查询性能。下面具体分析  $N_1$  数据集大小对第一阶段搜索效率的影响。由第一阶段搜索算法可知, 在  $N_1$  集合中插入一个向量  $p_i$  需要下列额外的运算步骤: 首先, 计算其距离上界  $u_i$ , 然后  $u_i$  与第  $k$  小的距离  $knn\_u[k]d$  进行比较, 如果  $u_i < knn\_u[k]d$ , 则  $knn\_u[k]d$  更新为  $u_i$ , 并且将  $knn\_u$  重新排序; 其次, 需要将向量标记  $i$  和距离下界  $l_i$  插入到堆  $heap\_l$  中。减少  $N_1$  中特征向量数目, 可以一定程度上降低第一阶段搜索时间, 而减少  $N_2$  中特征向量

数目, 则可以降低随机 I/O 次数。

在相关反馈过程中, 由于特征权重的改变, 需要重新扫描整个近似文件, 这时每个轮次的查询时间基本上是相同的。由于  $N_1$  集合中近似向量的数目很少, 而且  $N_1 \supseteq N_2$ , 可以在内存中继续保存这些集合, 在新一轮的查询过程中加以利用, 以提高检索速度。

采用  $t$  表示反馈过程的轮次,  $t=1, 2, \dots, T$ , 用  $q_t$  和  $W_t$  表示第  $t$  次反馈过程后的计算所得的查询向量和相似矩阵, 分别用  $N_1^t$ 、 $N_2^t$  表示第  $t$  次查询过程后的候选向量集合和特征向量集合。

从  $N_1^t$  中取出所有近似向量, 根据改变后的  $q_t$  和  $W_t$ , 可以重新计算出  $N_1^t$  中  $k$  个最小的距离上界, 用数组  $u\_d'$  表示。  $N_2^t$  集合中包含具体的特征向量, 从  $N_2^t$  中取出所有特征向量, 根据  $W_t$  计算  $q_t$  这些向量的实际距离, 用数组  $d'$  保存  $k$  个最小的实际距离。可以证明, 采用  $u\_d'$  或  $d'$  作为初始的距离上界来进行第一阶段过滤, 不会造成错误遗漏 (崔江涛 2005)。

$d'$  虽然是精确距离, 但是可以加入到距离上界数组中对向量进行过滤。重新对向量近似方法进行分析, 由于第一阶段顺序访问近似向量, 只能依靠距离上界来对向量进行过滤, 采用查询向量与特征向量的精确距离来过滤数据肯定可以增强过滤效率, 但是会破坏近似向量文件的顺序访问特性。

在进行下一轮次的查询过程时, 从数组  $u\_d'$  和  $d'$  中选择最小的  $k$  个距离组成数组  $knn\_u$ , 需要注意的是, 每个特征向量计算所得的距离上界和精确距离中只能选择数值较小的一个, 这样才能保证下一轮次搜索结果的完备性。下面给出  $t+1$  轮次近邻搜索算法中的第一阶段过滤算法, 第二阶段过滤算法与 VA 方法相同。

第  $t$  轮查询的第一阶段过滤算法过程如下 (崔江涛等 2006a):

- (1) 根据第  $t$  轮查询后的  $N_1^t$  和  $N_2^t$  集合, 分别计算数组  $u\_d'$  和  $d'$ ;
- (2) 选择数组  $u\_d'$  和  $d'$  中最小的  $k$  个值, 组成数组  $knn\_u$ ;
- (3) 根据  $a_i$ 、 $q_t$  以及  $W_t$ , 计算  $u_i$  和  $l_i$ ;
- (4) 如果  $l_i > uDist[k]$ , 则排除向量  $p_i$ , 否则将  $p_i$  插入到候选集中, 并用  $u_i$  更新数组  $uDist$ 。

上述改进的  $k$  近邻搜索算法虽然是针对二次式距离提出, 但是它也可以用到核函数方法中。

## 7.9 高维索引技术评价准则

高维索引技术主要应用领域就是多媒体数据库中的信息查询, 为了进行相似性查询, 数据结构的复杂性一般较高, 而且不同应用也具有不同的相似度量方法。另外, 随着计算机软、硬件的迅猛发展和 Internet 应用, 多媒体数据库往往具有海量特性和动态特性, 进行相似性查询的时间代价往往较高。为了支持上述应用, 高维索引结构应该具有以下特性:

- (1) 动态构造, 支持数据的动态插入和删除;
- (2) 二级/三级存储管理, 尽管主存容量日益增大, 仍不能将一个大型的完整的数据库保存在内存中, 因此要考虑二级以及三级的存储管理;

- (3) 独立于输入数据及插入顺序, 支持各种多维数据以及任意的插入顺序;
- (4) 可增长性, 索引结构要能适应数据库大小的增长;
- (5) 支持尽量多的操作, 支持多种相似度度量方式;
- (6) 空间的有效性, 一个索引结构相对于其原数据应该是比较小的;
- (7) 并行性及可恢复性。

评价一个索引结构, 主要从其查询过程的 I/O 复杂度和 CPU 复杂度两方面来评价。在传统索引结构中, 由于访问数据量较少, 主要采用 I/O 次数 (数据页面访问数量) 来评价索引结构。但是在高维情况下, 由于不存在亚线性复杂性的索引方法, 访问数据量较大, 查询过程的 CPU 代价也会影响到查询性能。

I/O 代价主要采用 I/O 次数来度量, 在定义数据页面大小的情况下, 一次 I/O 是指对一个数据页面的访问。对于外存, 对数据页面进行顺序访问是最快的。文献 (Berchtold S. et al 2000) 中提出的数据页面访问模型中, 访问一个数据页面的时间包括寻找特定页面的时间  $t_{seek}$  和数据传输时间  $t_{xfer}$ , 采用顺序访问方式, 可以节省寻找页面的时间  $t_{seek}$ 。顺序访问的速度一般要比随机访问速度快 5~10 倍。另外, 在某些索引结构中, 需要从数据页面中寻找某个特定的数据, 文献 (董道国等 2005) 中将这样一次访问看成一次 I/O, 但是文献 (Chakrabarti K. et al 2000) 对大量数据访问进行验证, 认为在乐观情况下, 这样一次访问可以看成 1/2 次 I/O。这样, 对于不同索引结构的访问方式, 根据 I/O 次数可以给出相应的 I/O 代价。以  $num\_Index$  表示树型索引结构 I/O 次数,  $num\_Sequence$  表示顺序访问 I/O 次数,  $num\_Random$  表示随机访问向量的数目, 传统树型索引结构 I/O 代价等  $num\_Index$ , 顺序访问方式下 I/O 代价等于  $\frac{num\_Sequence}{c}$ ,  $5 \leq c \leq 10$ 。随机访问方式下 I/O 代价为  $\frac{num\_Random}{2}$ 。

CPU 运算代价可以通过三种基本方式来度量:

- (1) CPU 运行时间, 可以在相同硬件情况下横向比较各种算法的 CPU 运行时间;
- (2) 数据访问数量, 如果每个数据点的 CPU 运算时间相同, 通过搜索算法中访问的数据量可以间接的比较 CPU 运算代价;
- (3) 采用加 (减) 法、乘法和比较运算的次数度量。

## 参 考 文 献

- [1] Arya S., Mount D.M., Netanyahu N.S., et al, 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of ACM*, 48 (6): 891-923.
- [2] Beckmann N., Kriegel H.P., Schneider R., et al, 1990. The R\*-tree: An efficient and robust access method for points and rectangles. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 322-331.
- [3] Bentley J.L., 1975. Multidimensional binary search trees used for associative searching, *Communications of the ACM*, 18 (9): 509-517.
- [4] Berchtold S., Bohm C., Kriegel H.P., 1996. The X-tree: An index structure method for high-dimensional data. In *Proc. ICVLDB*, 28-39.
- [5] Berchtold S., Bohm C., Braunmuller D.A., et al, 1997a. Fast parallel similarity search in multimedia

- databases. In Proc. ACM SIGMOD Int. Conf. on Management of Data, Arizona, USA, 1-12.
- [6] Berchtold S., Bohm C., Keim D.A., et al, 1997b. A cost model for nearest neighbor search in high-dimensional data space. In Proc. ACM PODS Conf., Tucson, AZ, 78-86.
- [7] Berchtold S., Ertl K.B., Kriegel H.P., 1998a. The pyramid-technique: towards breaking the curse of dimensionality. In Proc. ACM SIGMOD, Seattle, WA, 142-153.
- [8] Berchtold S., Keim A., 1998b. High-dimensional index structure: database support for next decades's applications. Tutorial Notes: ACM SIGMOD-98 Conference on Management of Data, Seattle.
- [9] Berchtold S., Bohm C., Jagadish H.V., et al, 2000. Independent quantization: an index compression technique for high-dimensional data spaces. Proc. IEEE Data Engineering. San Diego, 577-588.
- [10] Berchtold S., Bohm C., Keim D.A., et al, 2001. On optimizing nearest neighbor queries in high-dimensional data spaces. Proc. Int. Conf. on Database Theory, (ICDT 2001) London, 2001. in: Lecture Notes in Computer Science, Vol. 1973, Springer, 435-449.
- [11] Beyer K.S., Goldstein J., Ramakrishnan R., 1999. When is 'nearest neighbor' meaningful? Proc. 7th Int. Conf. on Database Theory (ICDT'99) . in: Lecture Notes in Computer Science, Vol. 1540, Springer, 217-235.
- [12] Bohm C., Berchtold S., Keim D.A., 2001. Searching in High-Dimensional Spaces-Indexing Structures for Improving the Performance of Multimedia Databases. ACM Computing Surveys, 33 (3): 322-373.
- [13] Cha G.H., Zhu X., Petkovic D., et al, 2002a. An efficient indexing method for nearest neighbor searches in high-dimensional image databases. IEEE Trans. Multimedia, 4 (2): 76-87.
- [14] Cha G.H., Chung C.W., 2002b. The GC-Tree: a high-dimensional index structure for similarity search in image databases. IEEE Trans on Multimedia. 4 (2): 235-247.
- [15] Chakrabarti K., Mehrotra S., 2000. Local dimensionality reduction: a new approach to indexing high dimensional spaces. In Proc. 26th International Conference on Very Large Databases, 89-100.
- [16] Chaur H.H., Yong J.L., 2000. Fast Search Algorithm for Vector Quantization of Images Using Multiple Triangle Inequalities and Wavelet Transform. IEEE Trans on Image Processing, 9 (3): 321-328.
- [17] Chavez E., Navarro G., Baeza-Yates R., et al, 2001. Searching in metric spaces. ACM Computing Surveys, 33 (3): 273-321.
- [18] Chen Y., Zhou X., Huang T.S., 2001. One-class svm for learning in image retrieval. In: Proc of IEEE ICIP, Thessaloniki, Greece, 815-818.
- [19] Ciaccia P., Patella M., 2000. PAC nearest neighbor queries: Approximate and controlled search in high-dimensional and metric spaces. In Proceedings of the 16th International Conference on Data Engineering, San Diego, California USA, IEEE Computer Society.
- [20] 崔江涛. 高维索引技术中向量近似方法研究, 博士学位论文, 西安电子科技大学, 2005.
- [21] 崔江涛, 孙君顶, 周利华. 基于小波变换的多分辨率高维图像检索方法, 西安电子科技大学学报, 2005, 32 (3): 370-373.
- [22] 崔江涛, 孙君顶, 周利华. 基于相关反馈的高维图像检索方法. 西安电子科技大学学报, 2006a, 33 (1): 62-65.
- [23] 崔江涛, 孙君顶, 付少峰等. 二次式距离上基于 SVD 的高维图像索引方法. 中国图像图形学报, 2006b, 11 (4): 498-503.
- [24] Cui J.T., Zhou S.X., Sun J.D., 2007. Efficient high-dimensional indexing by sorting principal component,

- Pattern Recognition Letters, 28: 2412-2418.
- [25] Dantong Yu, 2001. Multidimensional indexing and management for large-scale databases. PhD dissertation. University of New York.
- [26] 董道国, 薛向阳, 罗航哉. 多维数据检索结构回顾. 计算机科学, 2002, 29 (3): 1-6.
- [27] 董道国, 梁刘红, 薛向阳. VAR- tree 一种新的高维数据索引结构. 计算机研究与发展, 2005, 42 (1): 10-17.
- [28] Du H.C., Sobolewski J.S., 1982. Disk allocation for Cartesian product files on multiple Disk systems, ACM TODS, Journal of Transactions on Database Systems, 82-101.
- [29] 冯玉才, 曹奎, 曹忠升. 一种支持快速相似检索的多维索引结构. 软件学报, 2002, 13(8): 1678-1685.
- [30] Egecioglu O., Ferhatosmanoglu H., 2000. Dynamic dimensionality reduction and similarity distance computation by inner product approximations. In Proc. Of the 9th ACM Int. Conf. on Information and knowledge management, McLean, Virginia, 219-226.
- [31] Egecioglu O., Ferhatosmanoglu H., Ogras U., 2004. Dimensionality reduction and similarity computation by inner-product approximations. IEEE Tran. On Knowledge and Data Engineering. 16 (6): 714-726.
- [32] Faloutsos C., Bhagwat P., 1993. Declustering using fractals, PDIS Journal of Parallel and Distributed Information Systems, 18-25.
- [33] Feredman J.H., 1994. Flexible metric nearest neighbor classification. Department of Statistics, Stanford University, Tech Rep: 113.
- [34] Ferhatosmanoglu H., Tuncel E., Agrawal D., 2000. Vector approximation based indexing for non-uniform high dimensional data sets. Proc. of the ACM Int'l Conf. on Information and Knowledge Management (CIKM2000). New York: ACM, 202-209.
- [35] Ferhatosmanoglu H., Tuncel E., Agrawal D., et al. Approximate nearest neighbor searching in multimedia databases. In Proc. 17th IEEE Int. Conf. Data Eng. (ICDE), 503-511.
- [36] Finkel R., Bently J.L., 1974. Quad trees: A data structure for retrieval of composite keys. Acta Inf., 4 (1): 1-9.
- [37] Francois D., Wertz V., Michel V., 2004. Open questions about similarity search in high-dimensional spaces. 23th Benelux Meeting on Systems and Control.
- [38] Fukunaga K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, 2nd edition.
- [39] Gaede V, Gunther O., 1998. Multidimensional Access Method, ACM Computing Surveys, 30 (2): 170-231.
- [40] Guttman A., 1984. R-trees: A dynamic index structure for spatial searching. The ACM SIGMOD Int'l Conf. on Management of Data, Boston, MA, 47-57.
- [41] Hafner J., Sawhney H.S., Equitz W., et al, 1995. Efficient Color Histogram Indexing for Quadratic Form Distance Functions. IEEE Trans. on Pattern Analysis and Machine Intelligence. 17 (6): 729-736.
- [42] 贺玲, 吴玲达, 蔡益朝. CBIR 中的索引技术综述. 小型微型计算机系统, 2006, 27 (1), 141-145.
- [43] Heisterkamp D., Peng J., Dai H., 2001. An adaptive quasiconformal kernel metric for image retrieval. In: Proc IEEE CVPR, Kauai Marriott, Hawaii, 236-243.
- [44] Heisterkamp D.R., Peng P., 2003. Kernel VA-Files for relevance feedback retrieval. In Proc. First ACM international workshop on Mulitmedia databases, 48-54.
- [45] Hjaltason G., Samet H., 1995. Ranking in spatial databases, In Proc. 4th Int. Symp. on Large Spatial

- Databases, Portland ME. 83-95.
- [46] Indyk P., Motwani R., 1998. Approximate nearest neighbors: Towards removing the curse of dimensionality. Proc. 30th ACM Symposium on Theory of Computing, New York, ACM, 604-613.
  - [47] Ishikawa Y., Subramanya R., Faloutsos C., 1998. Mindreader: Query databases through multiple examples. Proc of 24th VLDB Conference. New York, Morgan Kaufmann, 218-227.
  - [48] Kanth K.V.R., Agrawal D., Singh A., 1998. Dimensionality reduction for similarity searching in dynamic databases. In Proc. ACM SIGMOD Int. Conf. on Management of Data, Seattle, Washington, 166-176.
  - [49] Katayama N., Satoh S., 1997. The SR-tree, An index structure for high-dimensional nearest neighbor queries. In Proc. ACM SIGMOD Int. Conf. on Management of Data, Tucson, Arizona USA, 369-380.
  - [50] Kim M.H., Pramanik S., 1988. Optimal file distribution for partial match retrieval. In Proc. ACM SIGMOD Int. Conf. on Management of Data, 173-182.
  - [51] Li J.H., 2001. Efficient similarity search based on data distribution properties in high dimensions. Ph.D dissertation, Michigan State University.
  - [52] Lin K., Jagadish H., Faloutsos C., 1995. The TV-tree: An index structure for high-dimensional data. Journal of VLDB, 3: 517-542.
  - [53] 刘芳洁, 董道国, 薛向阳. 度量空间中检索高维索引结构回顾. 计算机科学, 2003, 30 (7): 64-68.
  - [54] Lu G.J., 2002. Techniques and Data Structures for Efficient Multimedia Retrieval Based on Similarity, IEEE Tran. on Multimedia, 4 (3): 372-384.
  - [55] 陆哲明. 矢量量化编码算法及应用研究. 博士学位论文. 哈尔滨工业大学, 2001.
  - [56] Manjunath B.S., Ma W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Trans on Pattern Analysis Machine Intelligence, 18 (8): 837-842.
  - [57] Papadopoulos A., Manolopoulos Y., 1998. Similarity query processing using disk arrays. In Proc. ACM SIGMOD Int. Conf. on Management of Data.
  - [58] Peng J., Banerjee B., Heisterkamp D.R., 2002. Kernel index for relevance feedback retrieval in large image databases. In 9th International Conference on Neural Information Processing.
  - [59] Pramanik S., Alexander S., Li J., 1999. An efficient searching algorithm for approximate nearest neighbor queries in high-dimensions. In ICMCS 1999, IEEE International Conference on Multimedia Computing and Systems, Florence, Italy. IEEE Computer Society.
  - [60] Robinson J.T., 1987. The K-D-B-tree: A search structure for large multidimensional dynamic indexes. In Proc. ACM SIGMOD Int. Conf. on Management of Data, 433-438.
  - [61] Roussopolos N., Kelley S., Vincent F., 1995. Nearest neighbor queries, In Proc. ACM SIGMOD Int. Conf. on Management of Data, 71-79.
  - [62] Rui Y., Huang T.S., Mehrotra S., 1998. Relevance feedback: a Powerful tool in interactive content-based image retrieval. IEEE Trans. On CSVT, 8 (5): 644-655.
  - [63] Rui Y., Huang T.S., 2000. Optimizing learning in image retrieval. IEEE Conf. On CVPR, South Carolina, USA, 236-243.
  - [64] Sagan H., 1994. Space-Filling curves. Springer-Verlag, Berlin/Heidelberg/New York.
  - [65] Sakurai Y., 2000. The A-tree: an index structure for high-dimensional spaces using relative approximation. In: Proc. of the 26<sup>th</sup> VLDB Conf. Cairo, Egypt.
  - [66] Sellis T., Kriegel H.P., Faloutsos C., 1987. The R<sup>+</sup>-tree: A dynamic index for multidimensional objects.

- In Proc. 13th Int. Conf. on Very Large Databases, Brighton, GB, 507-518.
- [67] 孙圣和, 陆哲明. 矢量量化技术及应用. 北京: 科学出版社, 2002.
- [68] Tong S., Chang E., 2001. Support vector machine active learning for image retrieval. Proc. of ACM Multimedia, Ottawa, Canada.
- [69] Tuncel E., Ferhatosmanoglu H., Rose K., 2002. VQ-index: an index structure for similarity searching in multimedia databases. In Proceedings of the 10th ACM International Conference on Multimedia 2002, Juan les Pins, France. ACM, 543-552.
- [70] Wang J.Z., Li J., 2002. Learning-based linguistic indexing of pictures with 2-D MHMMs. ACM conference on Multimedia.
- [71] 汪祖媛, 庄镇泉, 王熙法. 逐维聚类的相似度索引算法. 计算机研究与发展, 2004, 41(6): 1003-1009.
- [72] Weber R., Schek H.J., Blott S., 1998. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. Proc. 24th Int. Conf. VLDB, New York, IEEE, 194-205.
- [73] Weber R., Bohm K., 2000. Trading quality for time with nearest-neighbor search. In Proc. Int. Conf. on Extending Database Technology. Konstanz, Germany, 21-35.
- [74] White D.A., Jain R., 1996. Similarity indexing with the SS-tree, In Proc. 12<sup>th</sup> Int. Conf. on Data Engineering, New Orleans, LA,.
- [75] Wu P., Schek H., Chandrasekaran S., 2001. An adaptive index structure for high-dimensional similarity search. In Proc. of Advances in Multimedia Information Processing-PCM 2001, the 2<sup>nd</sup> IEEE Pacific Rim Conf. on Multimedia. Lecture Notes in Computer Science 2195, Berlin: Springer-Verlag, 71-77.
- [76] 严蔚敏, 吴伟民. 数据结构(第二版). 北京: 清华大学出版社, 1992.
- [77] 杨志荣, 李磊. 用 SOM 聚类实现多级高维点数据索引. 计算机研究与发展, 2003, 40(1): 100-106.
- [78] 叶航军. 面向大规模图像库的索引和检索机制研究, 博士学位论文, 清华大学, 2003.
- [79] 叶航军, 徐光佑. 基于矢量量化的快速图像检索. 软件学报, 2004, 15(5): 712-719.
- [80] Yu C. High-dimensional Indexing: Transformational Approaches to High-dimensional Range and Similarity Searches. Lecture Notes In Computer Science. Springer-Verlag Heidelberg. 2002.
- [81] 于林森, 张田文, 张开越. 图像检索中的相似性判别及索引方法综述. 小型微型计算机系统, 2007, 28(2), 356-360.
- [82] Zhang C., Chen T., 2002. An active learning framework for content-based information retrieval. IEEE Trans. On Multimedia, 4(2): 260-268.
- [83] 张明波, 陆锋, 申排伟, 程昌秀. R 树家族的演变和发展. 计算机学报, 2005, 28(3): 289-300.
- [84] 周项敏, 王国仁, 于戈. 度量空间中索引方法的研究. 计算机科学, 2002, 29(B): 265-267.
- [85] 周项敏, 王国仁. 基于关键维的高维空间划分策略. 软件学报, 2004, 15(9): 1361-1374.
- [86] 庄毅, 庄越挺, 吴飞. 一种基于编码的双距离树高维索引. 中国科学(E 辑), 2007, 37(12): 1491-1503.



## 附录 A 基于轮廓的图像检索系统

这里我们结合第 4 章中基于轮廓特征进行检索为例，设计了一个基于形状轮廓的图像检索系统。下面就系统的设计、组成模块与功能等予以说明。

### A.1 系统框架

一般的图像检索系统主要由两部分组成：图像特征提取子系统和图像检索子系统，如图 A.1 所示。图像生成子系统主要包括图像输入前的预处理、图像内容的特征提取以及将特征和部分描述信息入库。图像检索子系统主要完成图像的检索，包括浏览查询、图像特征的相似度匹配和图形用户接口等部分。本系统的开发平台为 Window XP，开发工具为 Visual C++6.0 和 SQL2000。

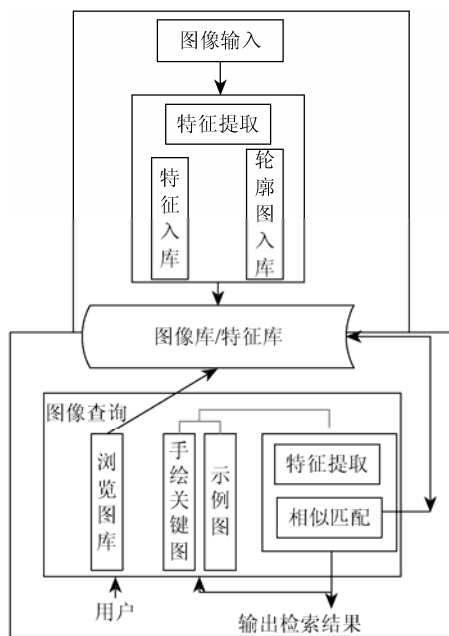


图 A.1 系统框架

### A.2 图像数据管理

基于内容的图像检索系统常常要涉及大量的数据，为了提高检索的效率，系统要对这些数据进行有效的管理。本系统建立的基于形状轮廓的图像检索系统所涉及的数据有：图像数据、提取的图像特征、用于分析形状的轮廓图等。

### (1) 图像数据

图像数据的存储方式有两种：一是把图像数据存入数据库，二是用文件管理的方式，图像数据仍存放于原来的文件系统中。前一种策略便于图像数据的管理，图像数据的安全性得到了保障，但同时给数据的存取带来了麻烦，而图像数据只是非结构数据，对其的操作完全依赖于应用程序本身，因此会影响图像的检索过程中对图像数据的操作效率。所有的图像数据都存入数据库的另一个弊端是数据库会变得非常庞大，不便于交流。第二种策略则便于图像数据的存取，同时便于数据的交流。因此本文采用第二种策略，即用以文件管理的方式对图像数据进行管理。

### (2) 图像特征数据

主要是轮廓特征，包括：对轮廓统计直方图、角点坐标和类型、轮廓点的相对位置分布和相关单元熵、各种不变矩向量等，这部分数据存放于SQL2000数据库中。

### (3) 轮廓图

为了方便研究和算法开发，我们将物体轮廓用有向链表表示，并存储在SQL2000数据库中。

### (4) 其他数据

其他的数据主要包括图像的路径信息、图像的标注信息等，与特征数据一起存放于数据库中。

## A.3 系统实现

### 1. 系统功能

系统主要功能包括以下几个方面：

#### (1) 图像入库操作

该系统的基本图像库包含MPEG-7标准形状检索库在内的共2075多幅图像，这些图像分为73个语义类。其中包括：鸟类、家禽、鱼类等动物，也包含有树木、花朵、钢笔等常见物体。库中图像均是单物体图像，方便了对形状的研究。系统支持用户继续向图像库中添加图像。

#### (2) 数据库浏览功能

包括按类别浏览以及随机浏览，同时用户可以根据浏览结果选择待查询图像。同时，浏览时可以同时观看用户在图像入库时所添加的备注信息；系统同时支持根据备注信息来进行基于文本的图像检索。

#### (3) 轮廓检索

系统中所提取的图像特征包括：轮廓直方图特征、改进直方图特征、轮廓矩、边界序列矩、相关单元熵、极半径不变矩、组合矩等特征；所支持的度量方式包括：欧氏距离、城区距离和综合加权三种方式。

#### (4) 基于绘制图的检索

用户可以根据需要绘制轮廓进行检索。

#### (5) 检索性能评价

支持基于“查准率”与“查全率”的评价方法，用户可以选择需要评价的算法，以及评价的一些参数，系统自动给出评价结果。

## (6) 其他功能

系统还支持轮廓图的自动角点提取。

## 2. 图像库构建

本系统主要研究分析物体轮廓，因此在预处理阶段，必须对图像进行分割，提取出物体轮廓。针对库中图像均是单物体的灰度图像，我们采用了canny边缘检测，然后运用形态学方法连接断开边缘的方法提取出物体轮廓。

我们的图像库对应三部分内容：其一为图像源库，即用来提取物体轮廓、图像的其他数据和结果显示时调用。其二为提取出的每幅图像对应的物体轮廓图所构成的轮廓图库，方便我们对轮廓描述算法的研究和特征的提取等。其三为存在SQL2000里的图像信息和特征向量。当将物体的轮廓均提取出来，并构建成轮廓库后，则可对每幅图像对应的轮廓图进行特征提取，最后将特征存入库中。

为了方便应用，本系统支持批量的图像预处理和提取特征，然后分别入库。图像入库流程如图A.2所示：

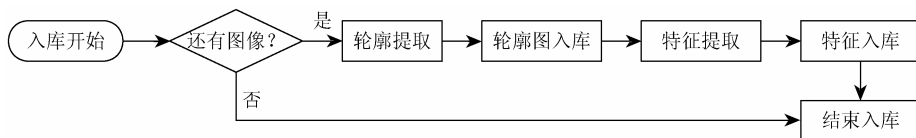


图 A.2 入库流程

图A.3为本系统中图像及特征入库的界面，用户可选择要入库的图像的路径，可以对选定图像的类别信息和其他备注进行标记，可以选择多个图像同时入库，也可以选择中列表头的序号前的选框对当前文件夹下所有图像进行入库。

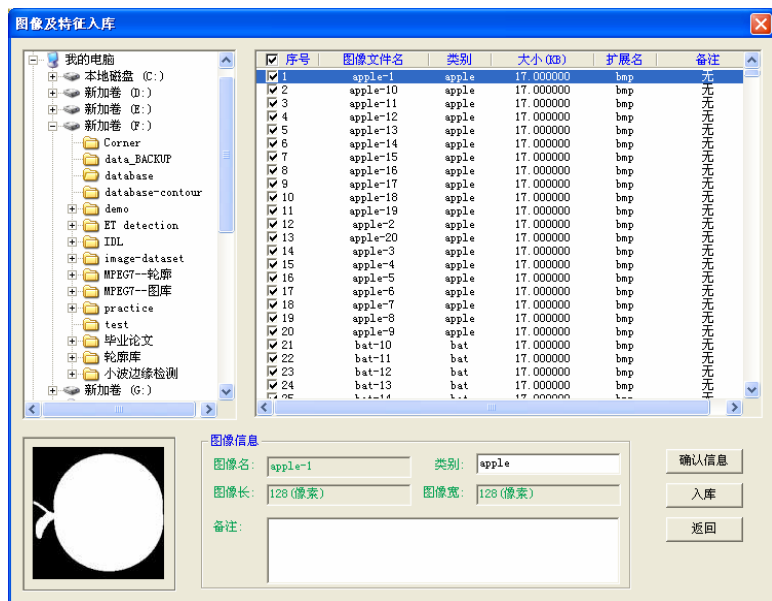


图 A.3 入库界面

### 3. 浏览查询模块

系统支持两种查询方式：浏览查询和示例查询。浏览查询是用户在现有图像库中浏览图像进行查询的方式，包括随机浏览和按类别浏览。图A.4给出了随机浏览时的用户界面，界面左边中间部分为浏览方式的选择和翻页操作，右边为随机的在图像库中访问的20个图像，左下部位为当前选中图像的预览区。图A.5给出了用户对现有图像库进行按类别浏览的系统界面，其中右部分为浏览显示区，当选择的类别的图像数目大于20时，可以通过左边的翻页操作进行浏览。

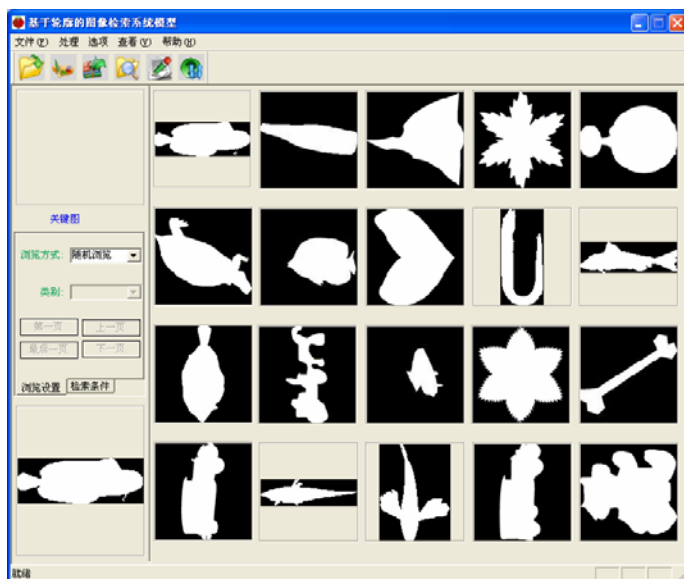


图 A.4 随机浏览图像库

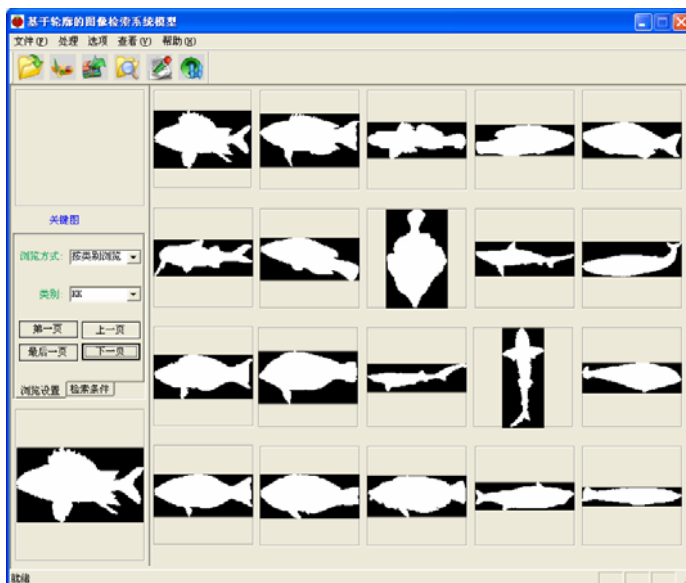


图 A.5 浏览图像库

浏览查询可以单独使用，也可以和样本查询配合使用，用于为示例查询选择查询示例图像。示例查询方式是由用户给出一个查询示例，然后系统根据查询示例，对图像库进行检索，按相似性大小排列出图像库与查询示例相似的图像。查询示例有两个来源：一是库中图像，即在浏览过程中选定图像；二是图像库以外的图像，可以是任何路径的图像。图 A.6 为用户选择图像库以外的图像时的界面，用户选择相应的文件后，可以在右侧看该文件的缩略图。

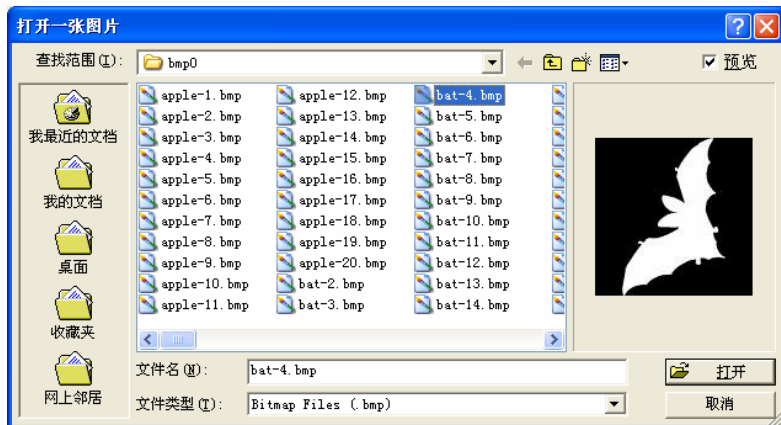


图 A.6 选择库外任意图

#### 4. 检索模块

图像检索流程如图 A.7 所示。图 A.8 为本系统支持的七种特征描述方法对应的一次检索效果图，在每个显示图像下方均有两个选择框，用户可以根据检索结果，评定是否符合检索要求，然后系统利用用户反馈的信息，确定权值以用于再次检索。图 A.9 为信息查看模块，在图像检索完成后，用户选择任一检索结果，可以查看该图像的详细信息（图像入库时，用户添加的备注信息）。

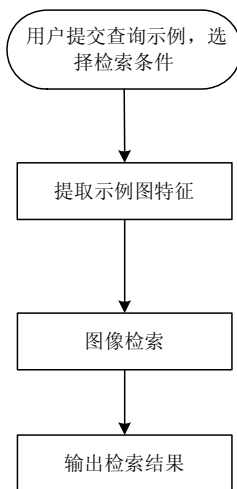
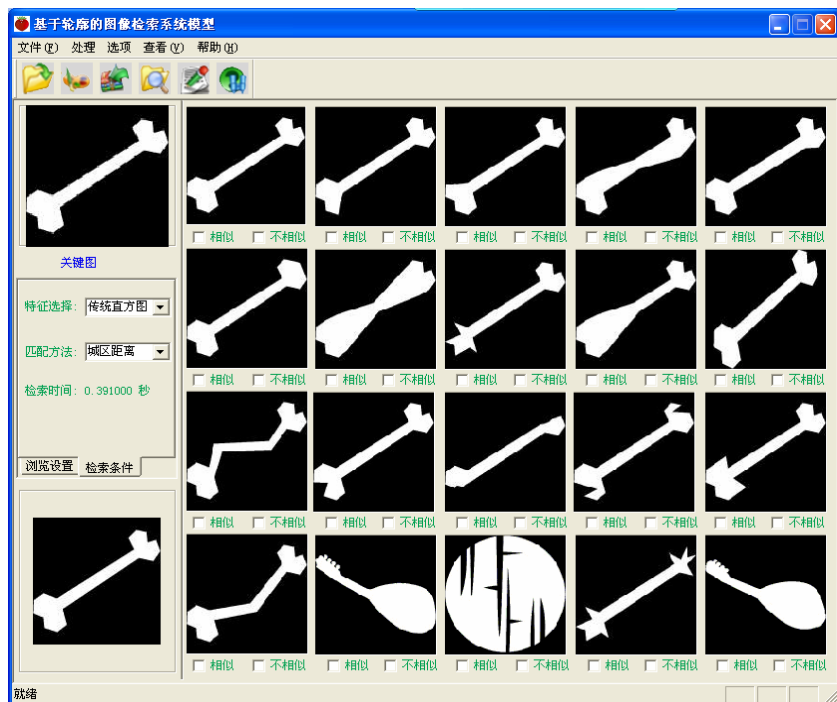
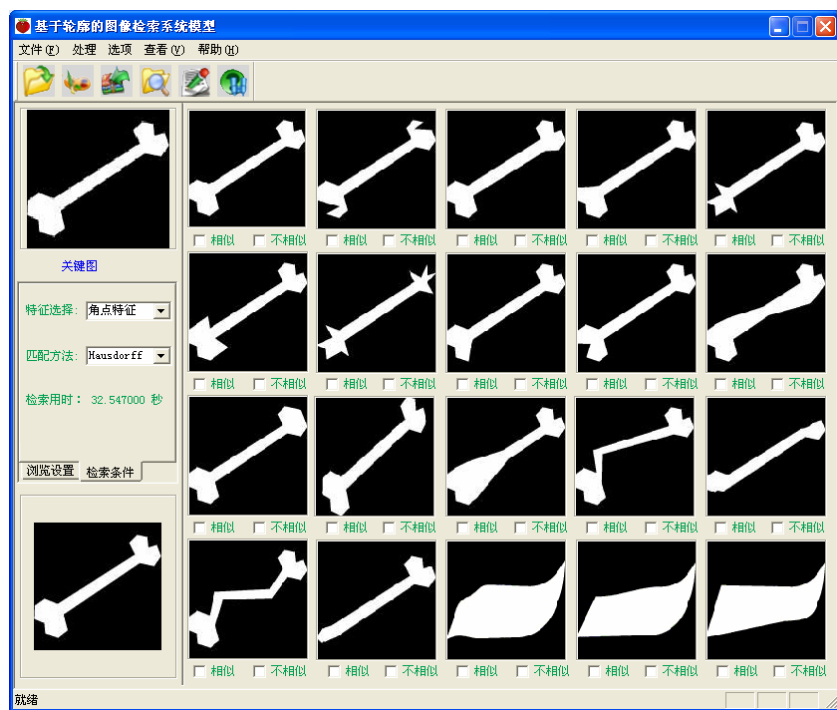


图 A.7 图像检索流程

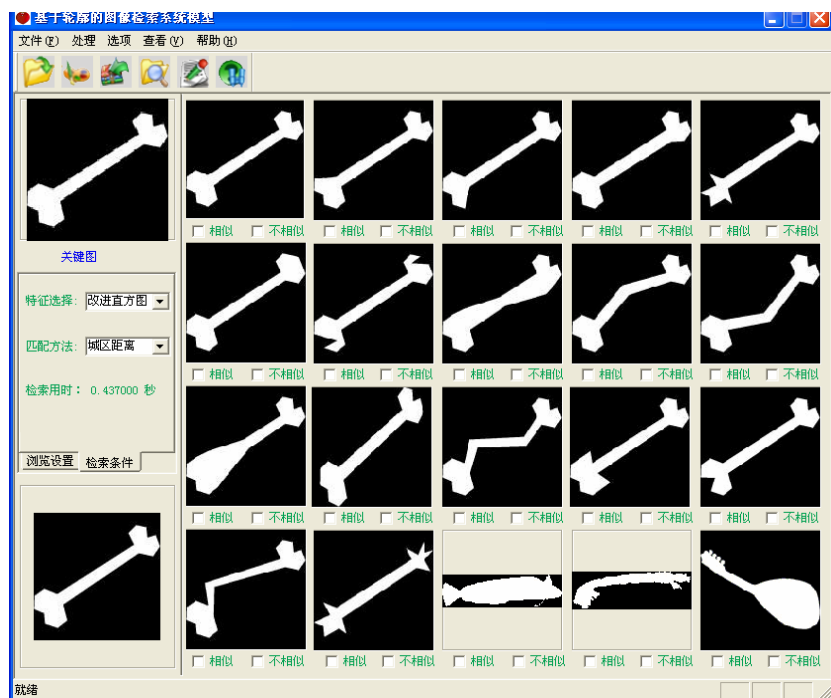


(a) 传统直方图方法

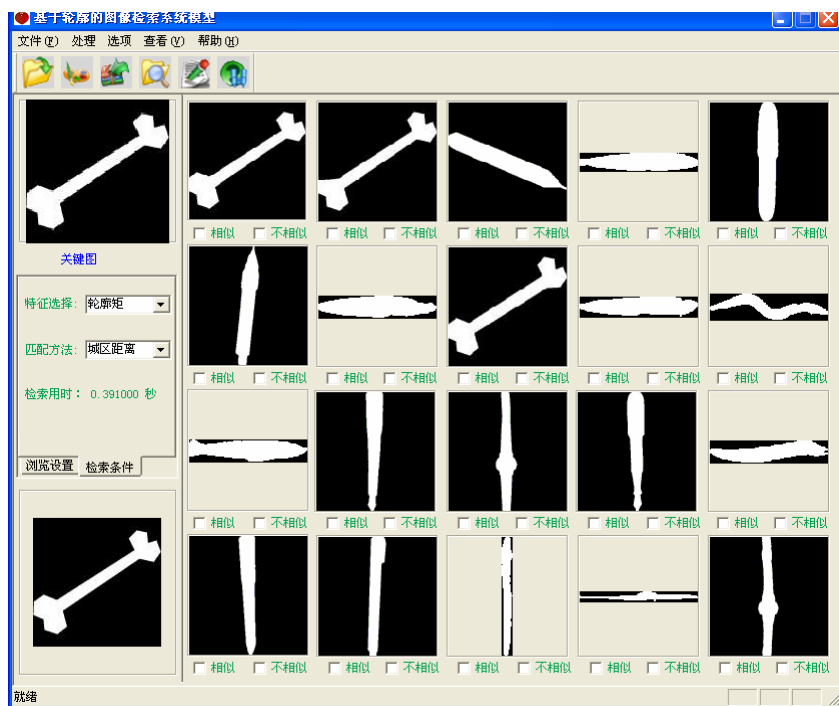


(b) Hausdorff 点匹配方法

图 A.8 7 种算法一次检索结果 (续)

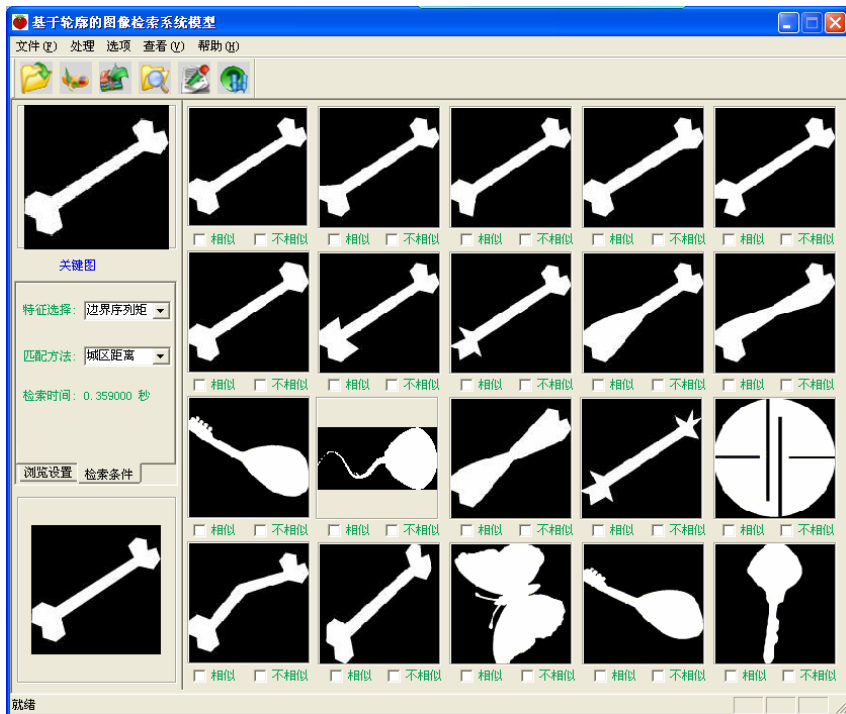


(c) 改进的直方图方法

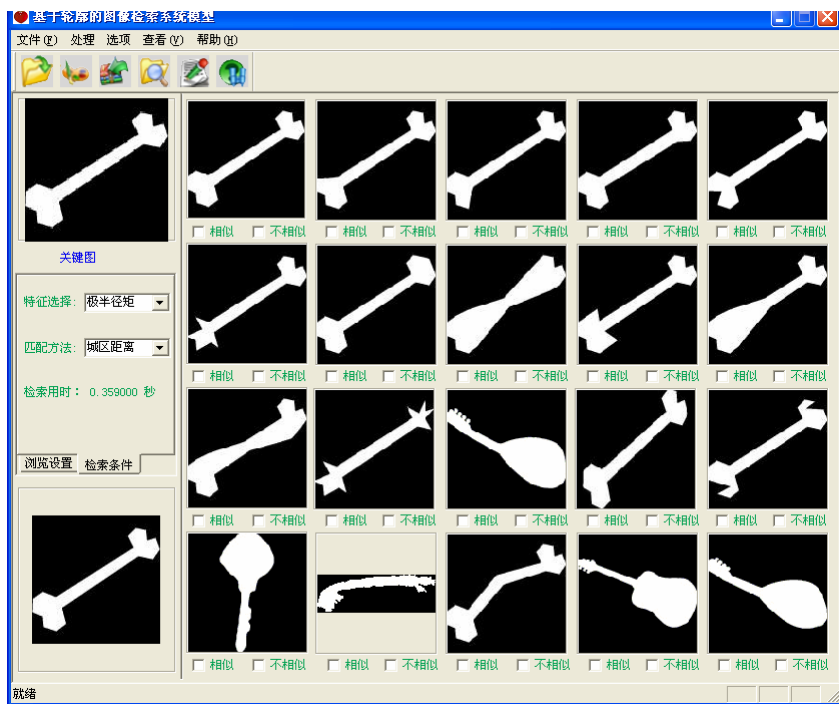


(d) 轮廓矩方法

图 A.8 7 种算法一次检索结果 (续)



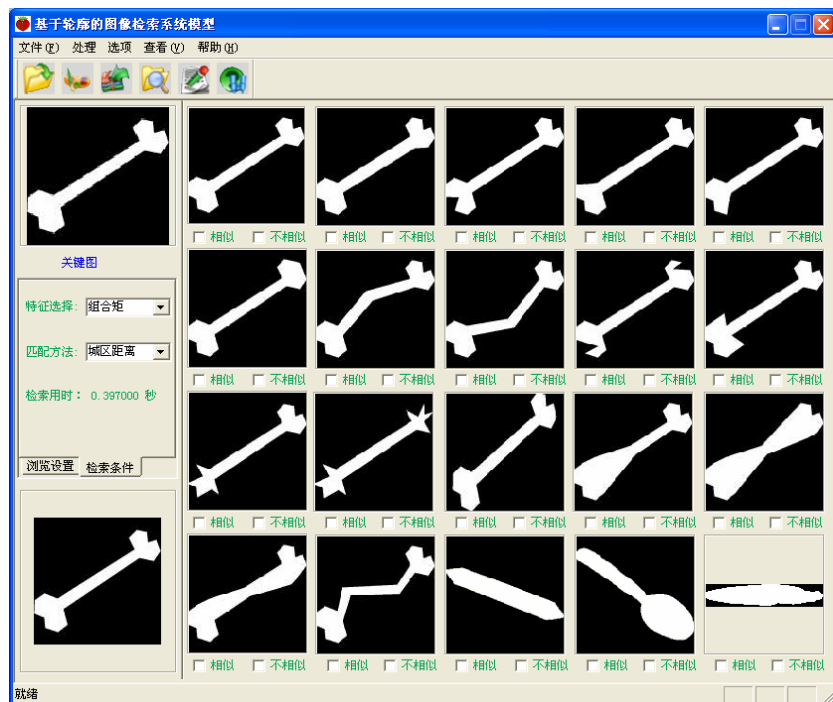
(e) 边界序列矩方法



(f) 极半径矩方法

图 A.8 7 种算法一次检索结果 (续)





(g) 组合矩方法

图 A.8 7 种算法一次检索结果 (续)

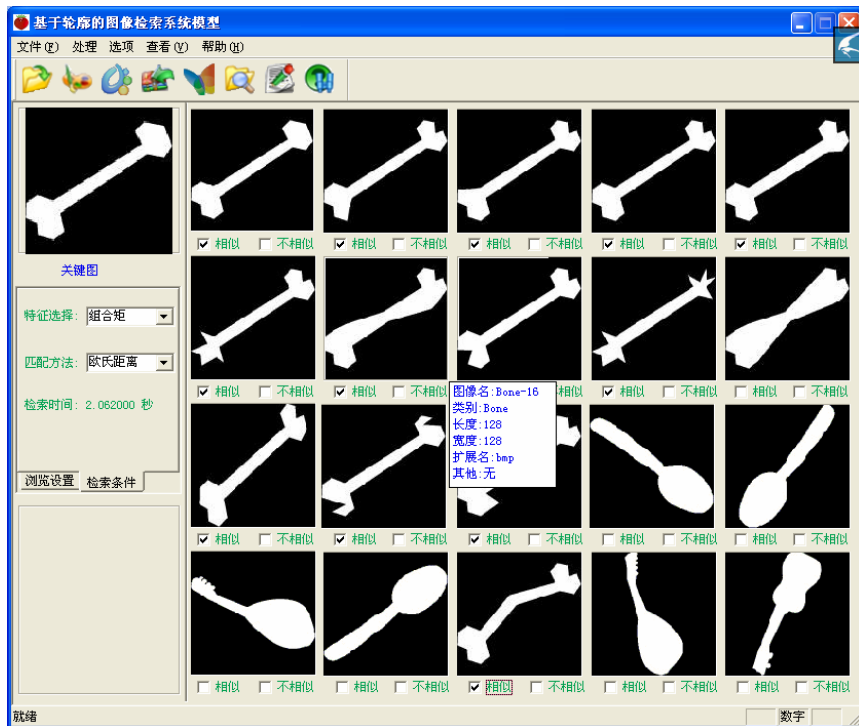


图 A.9 查看备注信息

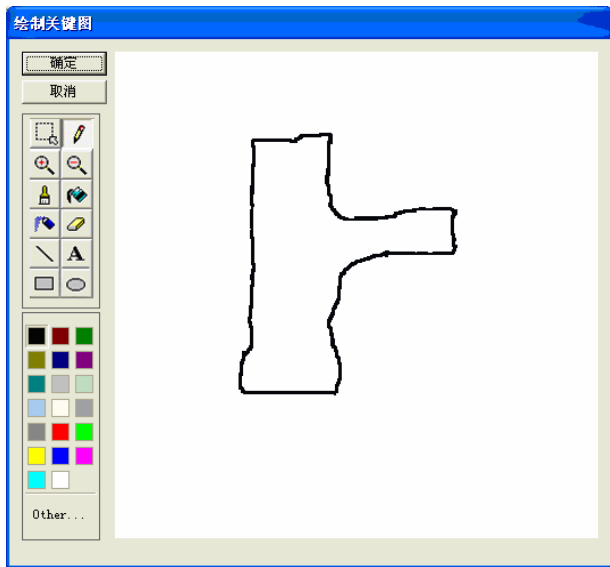


图 A.10 手绘轮廓

### 5. 基于绘制图的检索

本系统还支持用户根据需要，手绘制出示例物体的大致轮廓，然后以此轮廓图为例图进行查询。图A.10为用户使用本系统提供的绘制关键图功能绘制物体轮廓。

### 6. 性能评价模块

由于图像检索具有很强的主观性，因此，评价一个图像检索系统性能的优劣并不容易，往往需要结合多种性能评价指标。通常评价一个CBIR系统最重要的指标是查全率、查准率和响应时间。

本系统给出了三种统计查全率，查准率的方法：一是根据一次检索结果，统计结果和库中同类图像的数目，计算出查全率和查准率

；二是随机的从库中选择两幅图像，分别检索并计算出两组查全率查准率，最后取平均；三是随机的从库中取多张图像分别检索（我们选用的为20张），统计每次检索结果的查全率查准率；最后取平均作为评价算法的依据。用户可以选择一种或多种需要进行对比的算法。图A.11为性能评价的主界面。图A.12和图A.13给出了其中的2次评价结果。



图 A.11 性能评价界面

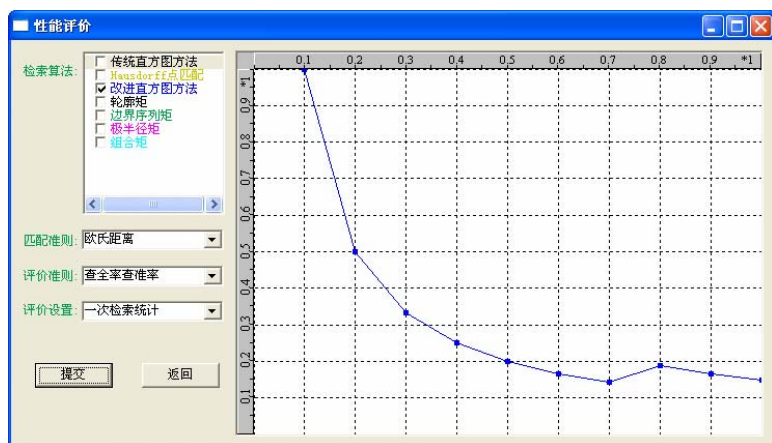


图 A.12 一次检索结果性能评价

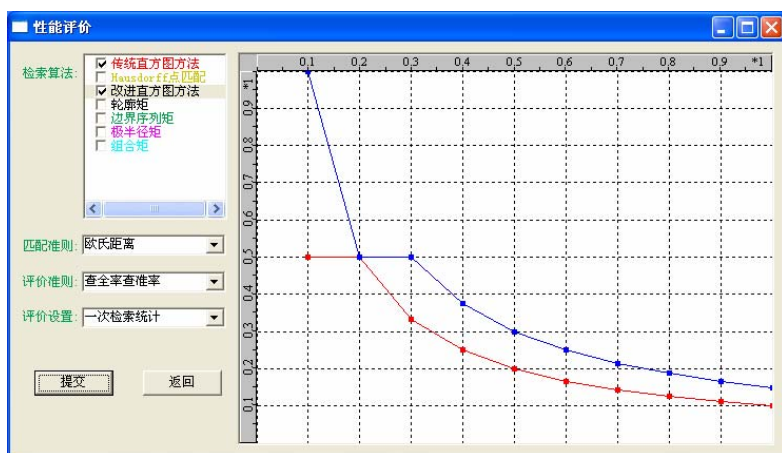


图 A.13 不同算法检索结果性能评价

## 7. 其他功能

系统支持轮廓角点的提取，如图 A.14 所示。

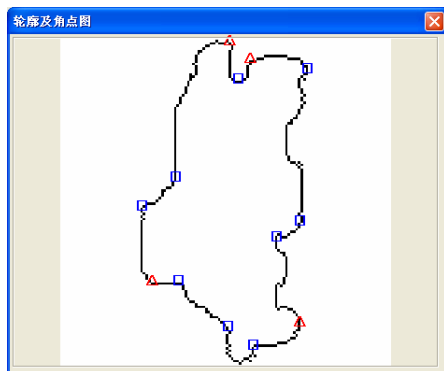


图 A.14 轮廓角点的提取

## 附录 B 系统中本文算法实现代码

### B.1 基于MCP的角点检测算法

```
%% MCP corner detection
clear all;
close all;
imagepath='F: \corner\airplane.gif';
I=imread ( imagepath );
BW = im2bw ( I, 0.5 );
[row, col]=size ( BW );
%% edge tracing
B=ones ( size ( BW ) +2 );
B ( 2: end-1, 2: end-1 )=BW;
[prow, pcol]= ( find ( B==0 ) );
points=contourtracing ( B, [prow ( 1 ) pcol ( 1 ) ] );
%%
curv_sum=ones ( length ( points ), 1 );
%curv_sum=zeros ( length ( points ), 1 );
for sig=1: 1: 5
    [gauss, width]=get_gauss ( sig ); %% Gauss smooth
    %% calculate curvature
    [K, contour]=curvature3 ( points, width, gauss );
    %% Find curvature local maxima as corner candidates
    extremum=[];
    newK=[];
    N=size ( K, 1 );
    n=0;
    Search=1;
    for j=1: N-1
        if ( abs ( K ( j+1 ) ) -abs ( K ( j ) ) ) *Search>0
            n=n+1;
            extremum ( n )=j;
            Search=-Search;
        end
    end
end
```

```

end
if mod ( size ( extremum, 2 ), 2 ) == 0
    n=n+1;
    extremum ( n ) =N;
end
curv_sum ( extremum ) =curv_sum ( extremum ) +K ( extremum );
id=1: length ( K );
id ( extremum ) =[];
newK=K ( extremum );
newK (: , 2 ) =extremum;
curv_sum ( id ) =curv_sum ( id ) .*K ( id );
subplot ( 5, 1, sig ), plot ( curv_sum, 'k' );
axis ( [0 length ( curv_sum ) -0.5 0.5] );

end
t1=0.1;
t2=-0.1;
curv_result=curv_sum;
curv_result (: , 2 ) =1: length ( curv_sum );
index=find ( 0<curv_sum (: , 1 ) &curv_sum (: , 1 ) <t1 );
curv_result ( index, : ) =[];
index2=find ( 0>curv_result (: , 1 ) &curv_result (: , 1 ) >t2 );
curv_result ( index2, : ) =[];
%% show in Original image
n=size ( curv_result, 1 );
c1=[];
c2=[];
for j=1: n
    if curv_result ( j ) >0
        c1=[c1; points ( curv_result ( j, 2 ), : ) ];
    elseif curv_result ( j ) <0
        c2=[c2; points ( curv_result ( j, 2 ), : ) ];
    end
end
figure, plot ( points (: , 2 ), points (: , 1 ), 'k' );
axis ( [0 col 0 row] );
axis ij;
hold on; plot ( c1 (: , 2 ), c1 (: , 1 ), 'bs' );
plot ( c2 (: , 2 ), c2 (: , 1 ), 'bv' ); hold off;

```

## B.2 组合矩算法

```

bool CContour::ChenMoment ( )
{
    Node central;
    int i=0; //循环变量
    double m00, m11, m20, m02, m30, m03, m12, m21; //中心矩
    double x0, y0; //临时变量
    double u20, u02, u11, u30, u03, u12, u21; //规范化
    double M[7]; //HU 不变量
    double t1, t2, t3, t4, t5; //临时变量
    int lBoundNum=m_BorderPoint.size ( );
    central=GetCenter ( lBoundNum );
    //计算 m00
    m00=lBoundNum;
    //计算 m11, m20, m02, m30, m03, m12, m21
    m11=m20=m02=m30=m03=m12=m21=0;
    for ( i=0; i<lBoundNum; i++ )
    {
        x0=m_BorderPoint[i].x - central.x;
        y0=m_BorderPoint[i].y - central.y;
        m11+=x0*y0;
        m20+=x0*x0;
        m02+=y0*y0;
        m30+=x0*x0*x0;
        m03+=y0*y0*y0;
        m12+=x0*y0*y0;
        m21+=x0*x0*y0;
    }
    //计算 u20, u02, m11
    u20=m20/ ( m00*m00*m00 );
    u02=m02/ ( m00*m00*m00 );
    u11=m11/ ( m00*m00*m00 );
    //计算 u30, u03, u12, u21
    u30=m30/ ( m00*m00*m00*m00 );
    u03=m03/ ( m00*m00*m00*m00 );
    u12=m12/ ( m00*m00*m00*m00 );
    u21=m21/ ( m00*m00*m00*m00 );
    //计算 7 个不变量

```

```

/* 中间变量*/
t1 = u20-u02;
t2 = u30-3*u12;
t3 = 3*u21-u03;
t4 = u30+u12;
t5 = u21+u03;
/*7 个不变量*/
M[0] = u20 + u02;
M[1] = t1*t1 + 4*u11*u11;
M[2] = t2*t2 + t3*t3;
M[3] = t4*t4 + t5*t5;
M[4] = t2*t4* ( t4*t4-3*t5*t5 ) + t3*t5* ( 3*t4*t4-t5*t5 );
M[5] = t1* ( t4*t4-t5*t5 ) + 4*u11*t4*t5;
M[6] = t3*t4* ( t4*t4-3*t5*t5 ) - t2*t5* ( 3*t4*t4-t5*t5 );
///记录提取特征
for ( i=0; i<7; i++ )
{
    m_chenMoment[i]=M[i];
}
return true;
}

bool CContour::CombMoment ( )
{
    int lBoundNum=m_BorderPoint.size ( );
    Node center=GetCenter ( lBoundNum );
    double x0, y0, temp; //临时变量
    double m1, M2, M3, M4; //中心矩
    double DN[3]; //3 个不变矩
    m1=M2=M3=M4=0;
    for ( int i=0; i<lBoundNum; i++ )
    { /*计算 m1*/
        x0=m_BorderPoint[i].x-center.x;
        y0=m_BorderPoint[i].y-center.y;
        m1+=sqrt ( x0*x0+y0*y0 );
    }
    m1=m1/lBoundNum;
    /*计算 M2, M3, M4*/
    for ( i=0; i<lBoundNum; i++ )
    {

```

```
        x0=m_BorderPoint[i].x-center.x;
        y0=m_BorderPoint[i].y-center.y;
        temp=sqrt ( x0*x0+y0*y0 ) -m1;
        M2+=temp*temp;
        M3+=temp*temp*temp;
        M4+=temp*temp*temp*temp;
    }
    M2=M2/lBoundNum;
    M3=M3/lBoundNum;
    M4=M4/lBoundNum;
    /*计算 4 个边界序列矩*/
    DN[0] = M2/ ( m1*m1 );
    DN[1] = M3/ ( m1*m1*m1 );
    DN[2] = M4/ ( m1*m1*m1*m1 );
    ///记录提取特征
    for ( i=0; i<3; i++ )
    {
        m_combMoment[i]=DN[i];
    }
    return true;
}
```



# 反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396; (010) 88258888

传    真：(010) 88254397

E-mail:    dbqq@phei.com.cn

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮    编：100036